nature

Análise | Publicados:08 de janeiro de 2020

Uma breve história da genética das doenças humanas

Melina Claussnitzer ,Judy H. Cho ,Rory Collins ,Nancy J. Cox ,Emmanouil T. Dermitzakis ,Matthew E. Hurles ,Sekar Kathiresan ,Eimear E. Kenny ,Cecília M. Lindgren ,Daniel G. MacArthur ,Kathryn N. Norte ,Sharon E.Plon ,Heidi L.Rehm ,Neil Risch ,Charles N. Rotimi ,Jay Shendure ,Nicole Soranzo eMark I. McCarthy

Natureza **577** , 179–189 (2020)

115 mil acessos | **319** citações | **652** Altmétrico | Métricas

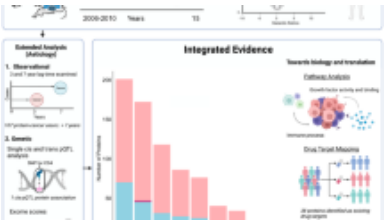
Abstrato

Um objetivo principal da genética humana é identificar variantes de sequências de DNA que influenciam características biomédicas, particularmente aquelas relacionadas ao início e progressão de doenças humanas. Ao longo dos últimos 25 anos, o progresso na concretização deste objectivo foi transformado pelos avanços na tecnologia, nos recursos genómicos fundamentais e nas ferramentas analíticas, e pelo acesso a grandes quantidades de dados de genótipos e fenótipos. As descobertas genéticas melhoraram substancialmente a nossa compreensão dos mecanismos responsáveis por muitas doenças raras e comuns e impulsionaram o desenvolvimento de novas estratégias preventivas e terapêuticas. A inovação médica centrar-se-á cada vez mais na prestação de cuidados adaptados aos padrões individuais de predisposição genética.

Conteúdo semelhante sendo visualizado por outras pessoas

Um catálogo profundo de variação na codificação de proteínas em 983.578 indivíduos

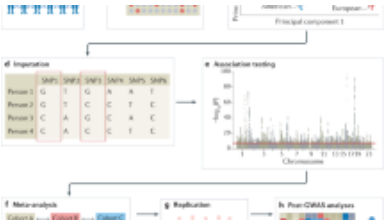
Artigo | 20 de maio de 2024



Identificação de fatores de risco proteômicos para câncer usando análises prospectivas e exoma d...

Artigo | Acesso aberto

15 de maio de 2024



Estudos de associação em todo o genoma

Artigo | 26 de agosto de 2021

Principal

Para quase todas as doenças humanas, a suscetibilidade individual é, até certo ponto, influenciada pela variação genética. Consequentemente, caracterizar a relação entre variação de sequência e predisposição para doenças fornece uma ferramenta poderosa para identificar processos fundamentais para a patogênese da doença e destacar novas estratégias para prevenção e tratamento.

Nos últimos 25 anos, avanços na tecnologia e nas abordagens analíticas, muitas vezes com base em grandes projetos comunitários - como aqueles que geraram a sequência do genoma humano [1] e elaboraram essa referência para capturar locais de variação genética [2,3,4,5,6] —permitiram a identificação de muitos dos genes e variantes que são causais de doenças raras e permitiram uma dissecação sistemática da base genética de características multifatoriais comuns. Há um impulso crescente por trás da aplicação deste conhecimento para impulsionar a inovação nos cuidados clínicos, mais obviamente através de desenvolvimentos na medicina de precisão. A medicina genómica, que anteriormente estava restrita a algumas indicações clínicas específicas, está prestes a tornar-se popular.

Esta revisão traça marcos recentes na história da genética das doenças humanas e oferece uma oportunidade para refletir sobre as lições aprendidas pela comunidade da genética humana. Concentramo-nos primeiro na divisão de longa data entre os

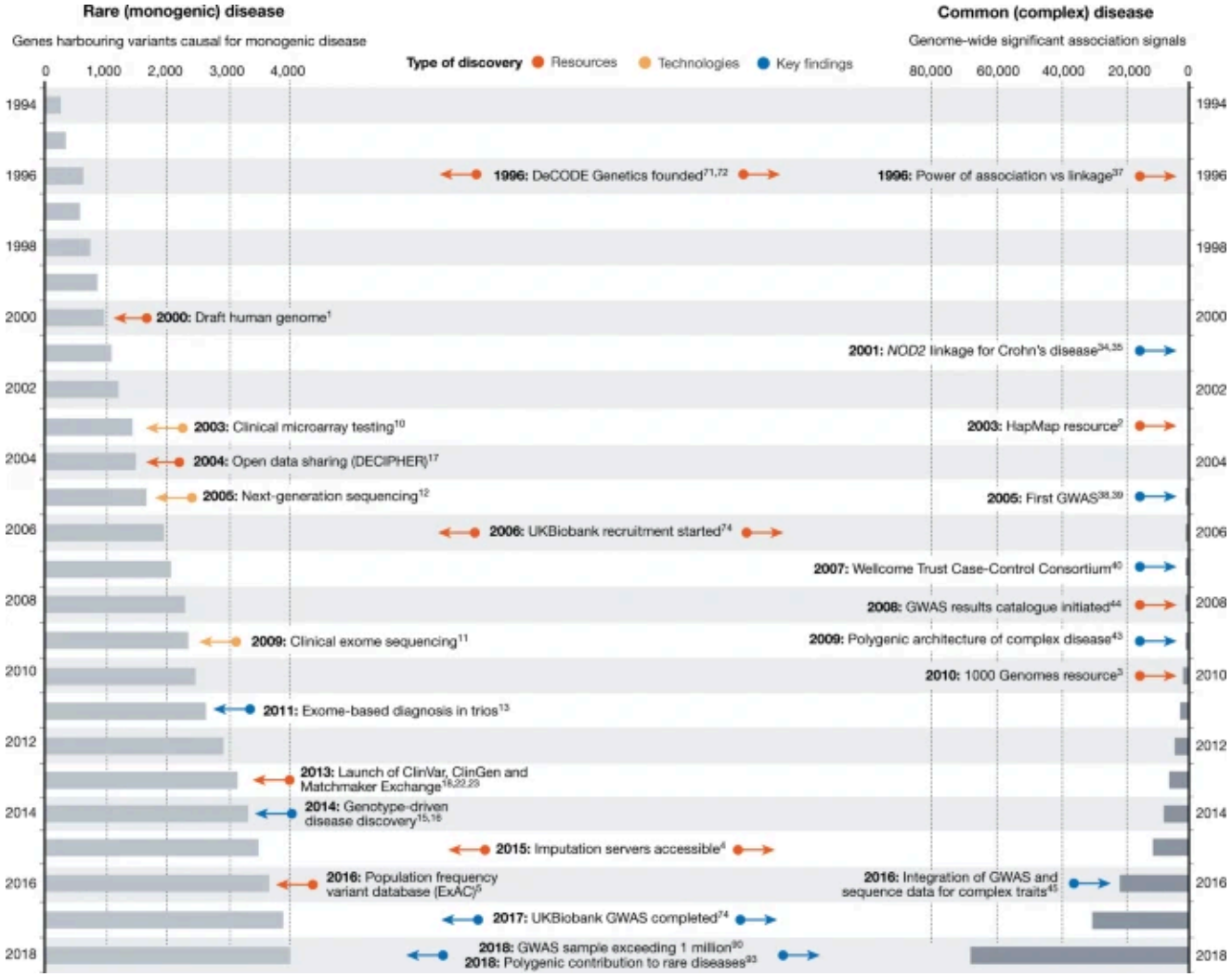
esforços de descoberta genética que visam variantes raras com grandes efeitos e aqueles que procuram alelos que influenciam a predisposição para doenças comuns. Descrevemos como esta divisão, com os seus ecos do debate centenário entre as visões mendeliana e biométrica da genética humana, obscureceu o espectro contínuo de alelos de risco de doença - em toda a gama de frequências e tamanhos de efeito - observados na população, e delineamos como as análises genómicas em grandes biobancos estão a transformar a investigação genética, permitindo uma perspectiva abrangente sobre as relações genótipo-fenótipo. Descrevemos como a expansão na escala e no escopo das estratégias para enumerar as consequências funcionais da variação genética está transformando a torrente de descobertas genéticas da última década em insights mecanicistas, e as maneiras pelas quais esse conhecimento sustenta cada vez mais os avanços nos cuidados clínicos. Finalmente, reflectimos sobre alguns dos desafios e oportunidades que enfrentam este campo, e os princípios que irão, durante a próxima década, impulsionar a aplicação da genética humana para melhorar a compreensão da saúde e da doença e maximizar os benefícios clínicos.

Doenças raras, variantes raras

Durante as décadas de 1980 e 1990, os esforços para mapear genes de doenças concentraram-se em doenças raras, monogênicas e sindrômicas e foram impulsionados principalmente por análises de ligação e mapeamento preciso em grandes linhagens multiplex. A localização dos sinais genéticos foi normalmente seguida pelo sequenciamento Sanger dos genes encontrados para mapear dentro do locus ligado para identificar alelos causadores de doenças. Avaliações de patogenicidade, baseadas na segregação de uma variante supostamente causal com a doença em múltiplas famílias e na evidência de que o genótipo de risco estava ausente em indivíduos saudáveis, eram normalmente seguidas por estudos funcionais confirmatórios em modelos celulares e animais. Esse caminho para a identificação genética foi trabalhoso; no entanto, até 2000, cerca de 1.000 das cerca de 7.000 doenças hereditárias de um único gene tinham sido caracterizadas, incluindo muitas com impacto biomédico substancial, como a doença de Huntington e a fibrose cística [7,8,9].

A conclusão do rascunho da sequência [1] do genoma humano reduziu muitos dos obstáculos ao mapeamento genético da doença e impulsionou um aumento de quatro vezes nos genes implicados como causais para distúrbios raros de um único gene (Fig. 1). A detecção de variação estrutural baseada em microarranjos [10] e o sequenciamento de todo o exoma e genoma [11,12] foram fundamentais, reforçados pela análise in silico e pela priorização das variantes genéticas descobertas. A crescente disponibilidade de conjuntos de dados de referência que catalogam a variação genética populacional em diversas origens étnicas apoiou uma inferência causal robusta [2,3,5,6] . Mais recentemente, a adopção de tecnologias de sequenciação de alto rendimento permitiu que toda a gama de variações genéticas causais, desde mutações únicas a grandes rearranjos estruturais, fosse identificada num único ensaio. Essas tecnologias se estenderam da pesquisa ao uso clínico, levando a diagnósticos mais precoces e rápidos de doenças genéticas.

Figura 1: Crescimento na descoberta de variação genética associada a doenças.



São mostrados os números cumulativos de genes que abrigam variantes causais de doenças e características monogênicas raras e de associações significativas de GWAS implicadas em doenças e características comuns e complexas. À esquerda, o advento de tecnologias de sequenciação de alto rendimento e a disponibilidade de genomas de referência de diversas populações apoiou um aumento de quatro vezes na descoberta de genes causadores de doenças raras entre 1999 e 2019. À direita, esforços internacionais como o Projeto Genoma Humano e o HapMap O projeto, combinado com GWAS e estudos de sequenciamento, apoiou a identificação de mais de 60.000 associações genéticas em milhares de doenças e características humanas. Centro, desenvolvimentos mais recentes trouxeram uma síntese das abordagens de variantes raras e comuns baseadas na combinação de análises informadas por sequência em grandes coortes. Os principais eventos que contribuem para esses temas são descritos na linha do tempo. GA4GH, Aliança Global para Genômica e Saúde [160] ; ExAC, Consórcio de Agregação Exome [5] .

A menor dependência de pedigrees multiplex em favor de coleções de casos afetados, muitas vezes com os pais [13] , provou ser decisiva na identificação de novos distúrbios dominantes, muitos dos quais eram anteriormente considerados

recessivos [14]. Cada vez mais, a descoberta de genes de doenças raras passou da caracterização genética de um pequeno número de indivíduos com apresentações clínicas semelhantes para o sequenciamento genômico de coortes maiores de pacientes fenotipicamente diversos. Esta abordagem baseada no genótipo revelou novos distúrbios associados a apresentações clínicas mais variáveis [15,16].

Uma abordagem mais sistemática à partilha de dados tem sido crítica, tanto para a caracterização de novas doenças como para a interpretação diagnóstica de potenciais alelos causais. O valor da partilha de dados genéticos e fenotípicos daqueles que se pensa serem portadores de doenças genéticas raras não diagnosticadas promoveu redes colaborativas globais (por exemplo, Matchmaker Exchange, DECIPHER e GeneMatcher) concebidas para combinar pacientes com variantes genéticas e/ou manifestações fenotípicas semelhantes, mesmo entre continentes [17,18,19]. As interações entre investigadores e famílias com doenças raras permitiram que estudos de história natural fossem conduzidos por grupos de apoio familiar posicionados para iniciar a recolha de dados de coortes de pacientes assim que um gene causal fosse descoberto [20].

A tradução clínica destas tecnologias beneficiou de uma série de recursos de informação, incluindo bases de dados abertas de genes associados a doenças raras (por exemplo, OMIM e ORPHANET) [21], variantes clinicamente interpretadas (por exemplo, ClinVar e ClinGen) [22,23] e registos de pacientes (por exemplo, DECIPHER e MyGene2 (https://mygene2.org/MyGene2)) [17]. O acesso a recursos que catalogam a variação genética entre populações (como o ExAC e o seu sucessor gnomAD) [5,6] permitiu a exclusão confiante de variantes genéticas demasiado comuns em dados a nível populacional para serem causas plausíveis de doenças genéticas raras e penetrantes de início precoce [24]. Estas análises reduziram a contaminação de bases de dados com variantes erroneamente interpretadas como causais de doenças e abordam a superestimação da penetrância da doença decorrente do foco histórico em pedigrees multiplex [25]. O melhor reconhecimento da penetrância variável de muitos alelos de doenças “monogênicas” revigorou os esforços para identificar os modificadores genéticos e ambientais responsáveis [26,27].

Embora tenham sido feitos enormes progressos na associação de genes específicos com doenças específicas, estabelecer o papel causal de variantes individuais dentro desses genes continua a ser problemático, e muitos pacientes com suspeita de doenças genéticas raras ficam sem um diagnóstico definitivo [28]. Mesmo para variantes com causalidade estabelecida, a penetrância muitas vezes não é clara. Resolver essas incertezas representa o desafio central para o campo. A agregação de dados de sequenciação de um grande número de casos afectados e de amostras de referência da população fornecerá a base de evidências necessária para uma interpretação robusta das variantes. Ensaios celulares in vitro altamente paralelizados que permitem a avaliação dos efeitos funcionais de todas as variantes em um gene associado à doença podem transformar a interpretação de novas variantes [29], embora o desenvolvimento de ensaios funcionais bem calibrados preditivos de patogenicidade para todos os genes da doença represente uma perspectiva assustadora. A exploração genômica funcional direta de tecidos acessíveis e relevantes para a doença de pacientes usando sequenciamento de RNA e ensaios de metilação de DNA [30,31] pode identificar variantes genéticas causais anteriormente enigmáticas, particularmente em regiões pouco exploradas fora dos genes codificadores de proteínas [32,33]. Os desenvolvimentos em cada uma destas áreas irão alargar a gama de variantes e genes para os quais a informação clínica diagnóstica e prognóstica pode ser fornecida aos pacientes e às suas famílias.

Doenças comuns, variantes comuns

Os esforços para aplicar a abordagem – análise de ligação em pedigrees multiplex – que tiveram tanto sucesso para as variantes de alta penetrância responsáveis pela doença mendeliana foram, com notáveis exceções [34,35,36], em grande parte malsucedidos para características comuns de início tardio com características mais complexas. etiologias multifatoriais, como asma, diabetes e depressão. O reconhecimento de que métodos baseados em associação, focados na detecção de diferenças relacionadas ao fenótipo nas frequências alélicas variantes, podem ter maior tração na identificação de alelos comuns menos penetrantes, redirecionou a atenção para a análise de amostras de caso-controle [37]. No entanto, os esforços iniciais visando variantes dentro dos genes “candidatos” foram atormentados por um poder inadequado, limites indevidamente liberais para declarar significância e pouca

atenção às fontes de preconceito e confusão, resultando em alegações exageradas e replicação falhada.

Esforços sistemáticos para caracterizar padrões de variação genômica em todo o genoma, inicialmente por meio do HapMap Consortium [2], mostraram-se catalíticos, demonstrando que a estrutura alélica do genoma foi segmentada em blocos de haplótipos, cada um contendo conjuntos de variantes correlacionadas. O reconhecimento de que esta configuração poderia apoiar pesquisas de associação em todo o genoma energizou a inovação tecnológica - na forma de matrizes de genotipagem massivamente paralelas - para tornar tais estudos possíveis (Fig. 1). As primeiras vitórias na degeneração macular aguda [38] e na doença inflamatória intestinal [39] foram encorajadoras, e o progresso em diversas frentes – expansão do tamanho do estudo, matrizes de genotipagem mais densas, novas estratégias de imputação, atenção a preconceitos e limiares de significância apropriados – proporcionou associações robustas numa série de doenças [40]. A maioria das variantes descobertas por estes primeiros estudos de associação genómica (GWAS) eram comuns, com efeitos mais subtis do que muitos tinham previsto. Foi formada uma série de consórcios específicos de características, cobrindo diversos fenótipos dicotômicos e quantitativos, para acelerar a descoberta genética através da agregação e meta-análise de dados de múltiplos estudos GWAS [41,42,43]. Foram identificadas muitas dezenas de milhares de associações robustas [44]. Recentemente, o aumento do acesso aos dados do exoma e da sequência do genoma completo, através da análise de associação direta [45,46] e da imputação [3,4], estendeu a descoberta para alelos raros e de baixa frequência, anteriormente inacessíveis ao GWAS.

Na década desde o primeiro GWAS, a compreensão da base genética das doenças humanas comuns foi transformada. A disparidade entre os efeitos observados das variantes identificadas pela primeira vez pelo GWAS e as estimativas da herdabilidade global das características (o enigma da «herdabilidade em falta») está agora amplamente resolvida [47]. As doenças comuns não são simplesmente agregações de condições mendelianas relacionadas: para a maioria das características complexas, a

predisposição genética é partilhada por milhares de variantes mais comuns, com efeitos individualmente modestos no risco populacional [41,43].

Embora a contribuição coletiva de alelos de risco raros e de baixa frequência para a variabilidade geral das características pareça modesta em comparação com aquela atribuível às variantes comuns [45,48], os alelos de risco raros detectados nos tamanhos de amostra atuais têm necessariamente grandes efeitos fenotípicos e são proporcionalmente mais propensos a serem codificação, aumentando seu valor para inferência biológica. As populações fundadoras (como as da Finlândia e da Islândia) forneceram vários exemplos de alelos de risco que de outra forma seriam raros, conduzidos a uma frequência local mais elevada através de deriva e/ou selecção [49,50,51,52]. Além disso, estudos em populações com altas taxas de consanguinidade tornam possível identificar indivíduos homozigotos para alelos de perda de função que de outra forma seriam raros, a base para um projeto de "nocaute humano" para investigar sistematicamente as consequências fenotípicas da ruptura genética em humanos [53,54].

Para a maioria das doenças, os esforços de agregação de GWAS em grande escala têm sido desproporcionalmente alimentados por informações provenientes de indivíduos de ascendência europeia [55]. Embora os padrões de predisposição genética pareçam amplamente semelhantes nos principais grupos populacionais e muitos alelos de risco comuns descobertos num grupo populacional sejam detectáveis noutros, as frequências dos alelos podem variar substancialmente; estender o GWAS e os estudos de sequenciamento para diversas populações certamente gerará uma rica colheita de novos alelos de risco.

As contribuições relativas das variantes comuns e raras indicam que, para muitas características, particularmente aquelas com início pós-reprodutivo, a selecção purificadora teve apenas um efeito limitado [45,56]. Para alguns alelos de risco, as características da selecção de equilíbrio reflectem o aumento da sobrevivência dos portadores, geralmente através da protecção contra doenças infecciosas. Isto inclui exemplos bem conhecidos de alelos mantidos com alta frequência em populações de ascendência africana [57,58].

Embora o extenso desequilíbrio de ligação nas populações humanas tenha sido essencial para a descoberta no GWAS, a alta correlação entre variantes adjacentes frustra o mapeamento das variantes específicas responsáveis por estas associações. O aumento do tamanho da amostra, a melhoria do acesso a dados transétnicos e painéis de referência de imputação mais representativos [3] fornecem um caminho para uma melhor resolução das variantes causais [59] e pistas para os mecanismos moleculares através dos quais elas operam. A interpretação funcional é mais fácil para variantes causais dentro de sequências de codificação; no entanto, as variantes de risco de doença mais comuns são mapeadas para sequências não codificantes e presume-se que influenciam a predisposição através de efeitos na regulação transcricional. Nestes casos, a inferência mecanicista depende da ligação dos sinais de associação aos seus alvos a jusante (ver abaixo). Para muitas características, existe uma clara convergência entre sinais de associação de variante comum e genes implicados em formas monogénicas da mesma doença, bem como enriquecimento de sinais GWAS em elementos reguladores especificamente activos em tipos de células consistentes com a biologia conhecida da doença [60,61]. Isto proporciona a garantia de que, mesmo que o número de sinais de associação para uma determinada doença prolifere, as associações genéticas descobertas irão coalescer em torno de processos moleculares e celulares com um papel central na patogénese [62,63].

É importante ressaltar que os sinais descobertos pelo GWAS revelaram muitos insights inesperados sobre a base biológica de doenças complexas. Exemplos incluem o papel do complemento na patogênese da degeneração macular aguda [38], da poda sináptica na esquizofrenia [64] e da autofagia na doença inflamatória intestinal [65]. Além disso, como a variação da sequência herdada é uma causa proeminente da variação fenotípica (mas o inverso não é verdadeiro), as variantes de risco identificadas pelo GWAS têm valor como instrumentos genéticos, mapeando relações causais entre características e inferindo contribuições feitas por biomarcadores circulantes e exposições ambientais a desenvolvimento de doenças [66].

Conforme descrito abaixo, os resultados do GWAS têm um impacto translacional crescente através da identificação de novos alvos terapêuticos [67], da priorização (e

desprioriização) dos existentes [68] e do desenvolvimento de pontuações poligênicas que quantificam o risco genético individual [69].

Mapas genótipo-fenótipo abrangentes

A divisão histórica da descoberta de genes de doenças em cadeias monogénicas e poligénicas surgiu do desenvolvimento e implementação de abordagens analíticas – ligação baseada na família e associação caso-controlo [37] – que são mais adequadas para detectar subconjuntos específicos de alelos causais. Isto obscureceu o verdadeiro estado da natureza, com os alelos de risco de doença distribuídos por um espectro contínuo de frequências e tamanhos de efeito. Além disso, a perspectiva específica de características e doenças da descoberta inicial de GWAS (principalmente dependente de estudos de caso-controle) estava mal equipada para investigar a contribuição de variantes genéticas para efeitos fenotípicos que estão aninhados ou espalhados pelas definições clássicas de doenças. Desenvolvimentos recentes permitiram uma perspectiva mais holística nas relações genótipo-fenótipo (Fig. 1).

Um grande avanço foi a crescente disponibilidade de grandes coortes prospectivas baseadas na população. Estes esforços de biobancos, pioneiros em estudos como a Coorte de Framingham [70] e os esforços do DeCODE na Islândia [71,72], abrangem agora um inventário crescente de coortes nacionais na América do Norte, Europa, Ásia e além [73,74,75,76]. O estudo UK Biobank, que incluiu 500 000 participantes de meia-idade, em grande parte saudáveis, tem sido particularmente influente, transformando a investigação genética humana, em parte através de políticas permissivas de partilha de dados que permitiram que vários grupos de investigação analisassem os dados [74]. Os esforços para disponibilizar dados clínicos incorporados em registos e registos de saúde eletrónicos para investigação [77,78] significam que os biobancos proporcionam cada vez mais acesso a uma vasta gama de dados demográficos, clínicos e de estilo de vida, capturados de forma harmonizada e sistemática a partir de grandes coleções, muitas vezes multiétnicas de indivíduos. Para milhões de participantes de biobancos, esta rica informação fenotípica foi combinada com dados genéticos de todo o genoma. Existem esforços nascentes para capturar fenótipos transcriptômicos, proteômicos e metabolômicos, embora estes

ainda não estejam em escala equivalente aos dados genéticos [79,80]. As análises de biobancos forneceram estimativas mais generalizáveis sobre a relevância dos fatores de risco genéticos no contexto dos efeitos separados e conjuntos de fatores não genéticos [81]. Cada vez mais, a integração com dados de saúde traz uma dimensão longitudinal à caracterização fenotípica, o que facilita análises da progressão da doença e do risco de doença ao longo da vida [82].

O rico escopo fenotípico dessas coortes permitiu que variantes de interesse fossem interrogadas quanto a associações em toda a gama de fenótipos disponíveis. Esses estudos de associação de todo o fenômeno (PheWAS) revelaram até que ponto muitas variantes têm efeitos pleiotrópicos em múltiplas características [83]. Algumas dessas relações são esperadas, como o impacto das variantes da obesidade no risco de esteatose hepática e diabetes tipo 2 [84] ou variantes que influenciam múltiplas condições autoimunes [85]. Outros conectam doenças e características de maneiras surpreendentes, destacando efeitos poligênicos e pleiotrópicos compartilhados e especificidade do tipo celular, e fornecendo insights sobre biologia compartilhada e mecanismos sobrepostos [86,87]. Estas descobertas informam a priorização de alvos terapêuticos, fornecendo pistas sobre potenciais efeitos colaterais no alvo e oportunidades para reaproveitamento de medicamentos [87,88,89].

O segundo facilitador da análise inclusiva e sistemática das relações genótipo-fenótipo tem sido o acesso aos dados da sequência do genoma completo. A escala da análise genética baseada em dados de sequência ainda está aquém da dos dados de genotipagem genómica (os maiores conjuntos de dados baseados em sequências têm um décimo do tamanho do maior GWAS [90,91,92]), embora as reduções nos custos de sequenciação estejam a diminuir o diferencial. A análise mais direta de dados de sequência de alto rendimento concentrou-se nas regiões de codificação. Estratégias para atribuir função variante e analisar conjuntamente conjuntos de variantes de efeito funcional semelhante permitiram testes agregados em nível de gene de rara associação funcional-variante que muitas vezes são mais bem alimentados do que testes de variante única [91,92]. No entanto, o principal benefício até o momento dos dados de sequência do genoma completo para a descoberta genética tem sido reforçar o acesso baseado em array a alelos de frequência mais baixa, seja

diretamente, através da sua inclusão em plataformas de genotipagem, ou indiretamente, através da imputação de referências baseadas em sequências. amostras [3,4].

Esses desenvolvimentos permitiram aos pesquisadores preencher a lacuna entre os domínios monogênico e poligênico, identificando variantes modificadoras comuns de fenótipos monogênicos que contribuem para a expressão variável de alelos raros e de grande efeito [26,93], e variantes raras e de baixa frequência que influenciam multifatoriais comuns. características [94,95]. Isto permite uma avaliação mais rigorosa da contribuição de variantes raras e comuns para a suscetibilidade das características [48] e apoia a enumeração de "séries alélicas" (conjuntos de alelos de frequência, tamanho de efeito e direção variados que perturbam o mesmo gene) essenciais para estudos do mecanismo da doença e otimização do alvo terapêutico [89,96]. Esses desenvolvimentos estão convergindo rapidamente para o destino final: uma matriz abrangente do efeito de todas as variantes genéticas observáveis na mais ampla gama possível de fenótipos biomédicos transversais e longitudinais. O sucesso neste esforço depende de uma harmonização cada vez maior e da integração dos resultados de estudos individuais através de investimentos sustentados na partilha de dados.

Adicionando função

Desde os primeiros mapas de ligação até à sequenciação do genoma completo de grandes coortes, a genética humana tem implementado abordagens cada vez mais sofisticadas e inerentemente sistemáticas para mapear os factores genéticos subjacentes às características e às doenças. No entanto, o progresso na determinação de como estas variantes influenciam a doença, através da interrogação sistemática dos seus efeitos funcionais nos processos moleculares, celulares e fisiológicos, tem sido muito mais lento.

Para doenças monogênicas, pelas quais os alelos responsáveis são tipicamente raros, penetrantes e codificadores, as abordagens genéticas têm sido geralmente necessárias e suficientes para implicar um gene como causal [28]. No entanto, à medida que os esforços para elucidar a base genética das doenças mendelianas progridem até à conclusão [97], os estudos funcionais continuam a ser importantes para

compreender os mecanismos pelos quais a variação disruptiva dentro de um gene causal leva a fenótipos de doenças. Ao contrário das doenças comuns, a clareza da causa dos distúrbios mendelianos geralmente simplifica a tarefa de gerar modelos (incluindo células humanas e organoides ou roedores) para conectar o genótipo ao fenótipo do organismo; estes levaram a muitos insights críticos sobre a biologia da saúde e da doença em humanos [98,99]. Além disso, para genes que abrigam variantes com consequências clinicamente acionáveis (como acontece com as mutações *BRCA1* e *BRCA2* que são causais do câncer de mama e de ovário de início precoce), estudos funcionais podem apoiar a interpretação translacional de novos alelos identificados por sequenciamento médico [29].

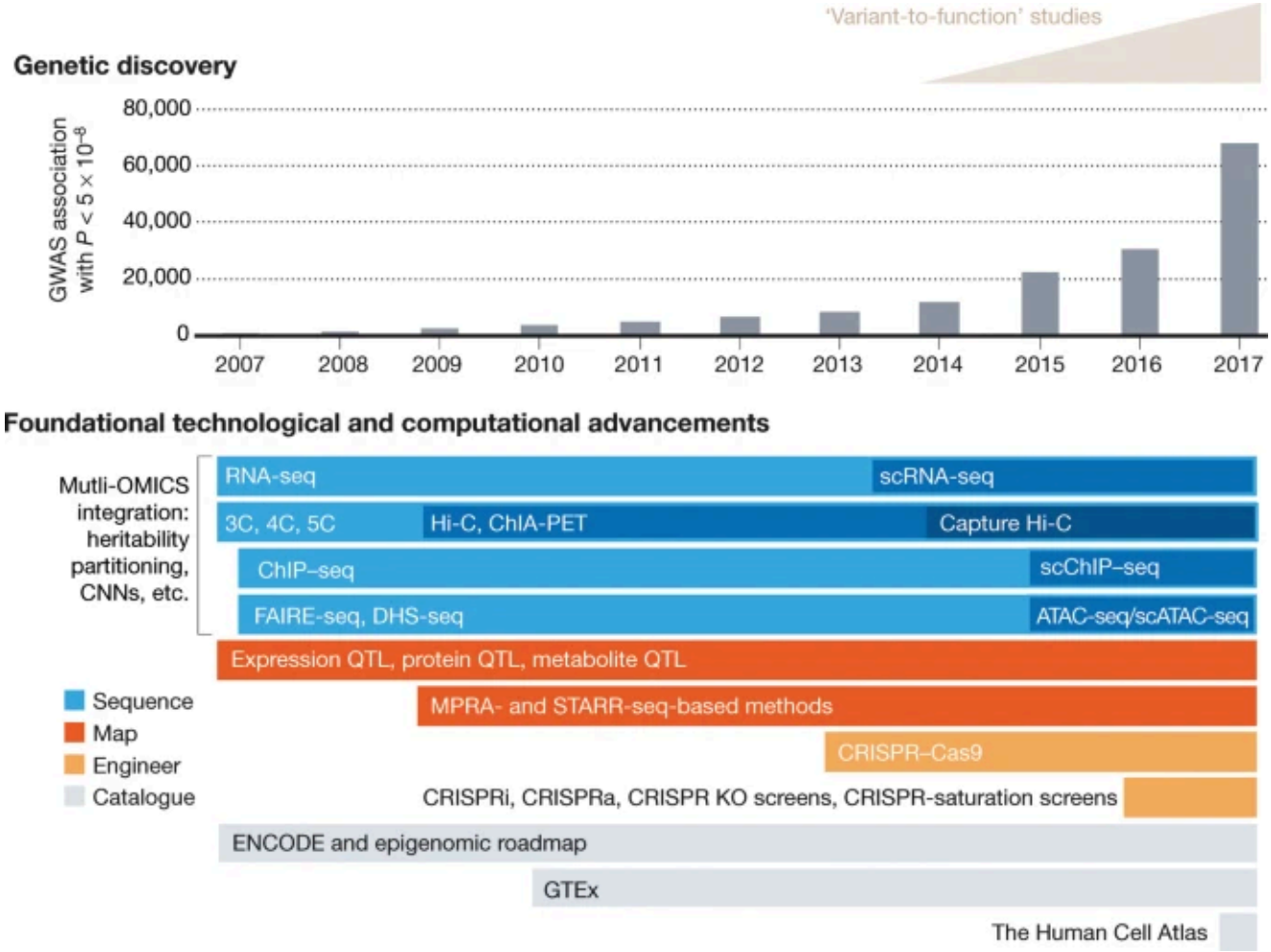
Para doenças comuns, os estudos funcionais têm um papel mais fundamental. Embora tenham sido descobertas dezenas de milhares de associações entre milhares de doenças e características humanas comuns [44], múltiplos factores frustraram os esforços para converter estes sinais genéticos em conhecimento sobre variantes, genes e mecanismos causais. Para as variantes comuns que estão subjacentes à maior parte do risco de doenças complexas, a resolução do mapeamento de associação é muitas vezes limitada pela estrutura haplotípica do genoma humano [2,3,4]. Além disso, a maioria das associações de GWAS são mapeadas para o genoma não codificante e, portanto, não possuem um endereço direto para o gene que medeia seus efeitos. A crescente apreciação do papel generalizado da pleiotropia complica a situação: muitas variantes identificadas pelo GWAS estão associadas a múltiplas características e exercem diversos efeitos em múltiplos tipos de células [100].

Até o momento, relativamente poucos estudos alcançaram o objetivo de conectar variantes causais de características complexas às funções moleculares e celulares que medeiam essa predisposição. Um sucesso inicial descreveu como variantes regulatórias que modulam a expressão *de SORT1* influenciam o colesterol de lipoproteína de baixa densidade e o risco de infarto do miocárdio [101]. Exemplos mais recentes concentraram-se na relação entre variantes associadas à obesidade intrônicas ao *FTO* , expressão alterada de *IRX3* e *IRX5* e função do adipócito [102] e do hipotálamo [103]. Descrições funcionais semelhantes foram relatadas para loci

individuais implicados na esquizofrenia [64], doença cardiovascular [104], diabetes tipo 2 [105] e doença de Alzheimer [106], entre outros.

Ao longo da última década, o desafio para a comunidade de genómica funcional tem sido converter este fluxo de trabalho de "um locus de cada vez" numa abordagem sistemática, multidimensional e integrativa, capaz de fornecer análises funcionais à escala do genoma que correspondam a todo o genoma. descoberta de variante (Fig. 2). No nível molecular, uma pedra angular tem sido a geração de catálogos de atividade funcional em todo o genoma. Por exemplo, os projetos ENCODE e Roadmap Epigenomics geraram mapas de modificações de histonas, ligação de fatores de transcrição, acessibilidade à cromatina, estrutura tridimensional do genoma e outras anotações regulatórias em centenas de tipos de células e tecidos [107,108]. Os padrões de sobreposição genómica entre estes dados e os resultados do GWAS permitem a inferência funcional de variantes de risco, fornecem pistas sobre os tipos de células específicos que conduzem a patogénese da doença [60,109] e aceleram os conhecimentos mecanísticos específicos do locus.

Figura 2: A descoberta genética é acompanhada de avanços nas tecnologias de genômica funcional.



Acima, o crescimento no número de loci genéticos associados pelo GWAS a características e doenças humanas (barras) e de estudos de variante para função (área abaixo da linha, sem escala). Abaixo, avanços tecnológicos e computacionais fundamentais na última década que permitiram (1) o desenvolvimento de catálogos sistemáticos de elementos funcionais em todo o genoma em vários tipos de células e tecidos (azul); (2) mapeamento de QTLs no contexto de expressão gênica, metabólitos, proteínas e elementos regulatórios (vermelho); (3) engenharia de genes, elementos genéticos e variação genética em escala crescente (laranja); e (4) levantamentos sistemáticos específicos de tecidos de elementos regulatórios e transcrição (cinza). scRNA-seq, análise de sequenciamento de RNA unicelular; ChIA-PET, análise de interação da cromatina por sequenciamento de tags paired-end; ChIP – seq, imunoprecipitação da cromatina seguida de sequenciamento; FAIRE-seq, isolamento de elementos reguladores assistido por formaldeído com sequenciamento; DHS-seq, sequenciamento de locais hipersensíveis à DNase I; ATAC-seq, ensaio para cromatina acessível por transposase usando sequenciamento; MPRA, ensaio repórter massivamente paralelo; STARR-seq, sequenciamento de região regulatória ativa autotranscritiva; CNN: redes neurais convolucionais. Para mais detalhes e literatura primária sobre muitos destes ensaios, consulte a ref. [173] .

Paralelamente, tem havido uma escalada de esforços para conectar variantes reguladoras associadas a características aos genes e processos que regulam em tipos de células relevantes para a doença de interesse [110,111]. Por exemplo, o consórcio GTEx (Genotype-Tissue Expression) mapeou milhares de loci de características quantitativas (QTLs) de expressão em centenas de indivíduos e dezenas de tecidos [112]. Pistas adicionais sobre as relações entre variantes reguladoras e seus genes efetores podem ser coletadas a partir de ensaios de proximidade de DNA (como Hi-C) e dados unicelulares [113] (Fig. 2). Programas como o HubMAP [114] e o Human Cell Atlas [115] foram criados para fornecer mapas de referência abrangentes e de alta resolução de tipos de células humanas individuais em diversos estágios de desenvolvimento, proporcionando novas oportunidades para entender como a variação genética regulatória resulta em fenótipos celulares e de organismos.

Os esforços para investigar as consequências clínicas da codificação de alelos com grandes efeitos fenotípicos (particularmente alelos nulos) em humanos [53,54] e em diversos modelos animais [116] representam estratégias poderosas para estender as análises funcionais ao nível do corpo inteiro. As conexões entre a variação genética e os dados proteômicos e metabolômicos circulantes fornecem ligações mecanísticas adicionais entre eventos celulares e a fisiologia de todo o corpo [79,80]. Esses esforços são paralelos às abordagens PheWAS [83], que, ao mapear efeitos variantes em toda a gama de características disponíveis em biobancos e EMRs, podem informar antecedentes para tipos de células e vias em loci individuais. É importante ressaltar que, embora os primeiros estudos normalmente vinculassem os alelos de risco do GWAS aos dados de um único ensaio funcional, o foco está cada vez mais na maximização do conhecimento biológico por meio da integração multidimensional de vários tipos de dados do genoma, usando abordagens como particionamento de herdabilidade [117], análises de enriquecimento funcional [60,109], integração da estrutura tridimensional do genoma [118] e redes neurais convolucionais profundas [119,120].

Embora as análises de QTL possam implicar um haplótipo num fenótipo molecular, celular ou de organismo, elas são, isoladamente, insuficientes para definir as variantes causais específicas responsáveis. Para resolver isso, houve um rápido

amadurecimento de tecnologias, como ensaios de repórter massivamente paralelos [121,122,123] e edição de genoma CRISPR, para apoiar a caracterização funcional de perturbações de sequência alvo em escala. Variações nesses métodos permitem a avaliação funcional de genes (por meio de telas knockout [124]), elementos reguladores (usando interferência CRISPR e telas de ativação CRISPR [125,126]) e variantes genéticas (editores de base [127]) em escala e resolução crescentes [29]. Combinados com leituras complexas - incluindo imagens de alto conteúdo [128] e transcriptômica e epigenômica unicelulares [129,130] - esses métodos podem gerar dados empíricos "verdadeiros", apoiando o desenvolvimento de modelos in silico para prever variantes causais, transcrições efetoras [126] e efeitos celulares . No devido tempo, tais modelos deverão reduzir a necessidade de caracterização experimental exaustiva da função para todas as variantes em todos os tipos de células.

O objetivo de tais esforços é enumerar a cascata de eventos moleculares que fundamentam as associações genótipo-fenótipo observadas usando sistemas celulares fisiologicamente relevantes (de células primárias a organoides e designs de 'órgão em chip') e ensaios de corpo inteiro apropriados para a doença de interesse. Coletivamente, estratégias que oferecem avaliação funcional em larga escala de variantes e genes de interesse reduzirão (mas provavelmente não eliminarão) o esforço intensivo necessário para a validação da 'milha final' dos mecanismos da doença em sistemas dedicados, acelerando assim a aplicação translacional a jusante.

Implementação clínica

A genética médica, aplicada às doenças raras, tem sido caracterizada pela rápida aplicação na clínica das tecnologias genómicas transformadoras que impulsionaram as descobertas iniciais da investigação. Existem agora testes genéticos direcionados para quase todas as apresentações clínicas atribuíveis a alelos de grande impacto, juntamente com ensaios mais extensos de sequenciação do genoma que, quando necessário, permitem a interrogação de uma lista mais longa de genes relevantes. Testes genéticos para indivíduos sintomáticos e parentes em risco ocorrem rotineiramente em muitas especialidades médicas. Paralelamente, a utilização de testes de cancro somático aumentou à medida que terapias direcionadas para

eventos mutacionais específicos entraram na prática clínica (estes desenvolvimentos são revistos noutro local [131,132]).

Para pacientes com sintomas que indicam uma provável etiologia monogênica (como degeneração da retina, perda auditiva ou cardiomiopatia), os painéis direcionados são normalmente a plataforma de escolha [133], embora sejam cada vez mais realizados em uma estrutura de sequência mais extensa. Para fenótipos mais complexos – aqueles sem uma correspondência clara com uma síndrome específica, como distúrbios do neurodesenvolvimento e múltiplas anomalias congênitas – os testes gravitaram em direção à implantação precoce de exoma e plataformas de sequenciamento de genoma que oferecem resolução rápida do que historicamente tem sido um diagnóstico traumático. odisseia [15,134]. O poder do diagnóstico genômico é especialmente claro para aqueles que apresentam distúrbios monogênicos do desenvolvimento neurológico e bebês gravemente doentes [135,136]. O sequenciamento do trio pai-filho pode detectar variação de novo em distúrbios dominantes e variantes raras bialélicas de fase na doença recessiva [13].

A transição de testes genéticos direcionados para o sequenciamento genômico permite reanálises recursivas, incluindo a reinterpretação de sequências individuais com base em descobertas subsequentes sobre alelos causais de doenças e suas consequências fenotípicas [137]. No entanto, são necessários diagnósticos moleculares melhorados para garantir a detecção confiável de um subconjunto de doenças genéticas, incluindo aquelas decorrentes de repetições triplas e rearranjos complexos [138]. O sequenciamento profundo de tecidos afetados para variantes em mosaico e o uso de sequenciamento de RNA para detectar variantes não codificantes que provocam doenças de início precoce (por exemplo, através de efeitos no splicing) representam novas frentes para o diagnóstico clínico [30].

Outros exemplos da rápida adoção de novas tecnologias genómicas incluem testes pré-natais não invasivos (mais de dez milhões de testes até 2018 em vários países [139,140,141]) e a utilização de painéis de portadores recessivos para casais que planeiam engravidar. A triagem neonatal é agora universal em muitos países, embora esteja limitada a doenças que combinam detecção de alto rendimento e baixo custo com

intervenções precoces eficazes (como restrições dietéticas ou reposição enzimática) [142]. O diagnóstico genético também é cada vez mais aplicado à triagem neonatal como um teste reflexo após um teste de triagem anormal (por exemplo, metabólico) [143]. Durante a próxima década, o repertório de doenças detectadas pelo rastreio neonatal e pelos testes pré-natais deverá expandir-se acentuadamente. Embora os testes pré-natais possam ser mais eficazes para evitar doenças, as questões éticas associadas são mais complexas [144].

Embora os testes genéticos para doenças raras e cancro tenham explodido, tem havido uma utilização mais limitada da informação genética noutros aspectos dos cuidados de saúde. Por exemplo, apesar de vários exemplos de marcadores genéticos clinicamente importantes relacionados com a eficácia dos medicamentos e o perfil de efeitos secundários [145], a implementação da farmacogenética tem sido dificultada por uma série de factores, incluindo a falta de apoio à decisão clínica em sistemas médicos electrónicos para orientar a escolha do medicamento ou dosagem pelo médico. Isto foi agravado por desafios nos testes de diagnóstico: estruturas complexas de haplótipos e variantes estruturais em alguns locais-chave do metabolismo de medicamentos necessitam de sequenciação do genoma ou painéis direcionados específicos para detectar todas as variantes clinicamente relevantes.

Para doenças comuns, a atenção translacional está atualmente focada no potencial clínico dos escores de risco poligênicos. O desenvolvimento de pontuações poligénicas robustas para várias doenças comuns foi catalisado por estimativas mais precisas do efeito por variante a partir de conjuntos de dados GWAS maiores, algoritmos melhorados para combinar informações através de milhões de polimorfismos de nucleótido único e biobancos em grande escala que suportam a validação de pontuações [69,146,147]. Por exemplo, uma pontuação poligénica genómica para ataque cardíaco, incorporando 6,6 milhões de variantes, indica que 5% dos indivíduos de ascendência europeia têm um risco de eventos cardíacos futuros equivalente ao observado naqueles com formas monogénicas menos frequentes de hipercolesterolemia [69]. Cada vez mais, a mudança da genotipagem baseada em arranjos para a análise baseada em sequências está facilitando a previsão de risco, que integra informações de alelos raros e de grande efeito com informações

de escores poligênicos [93]. Ao melhorar a captura do risco genético, especialmente em populações não europeias, e ao integrar dados ambientais e de biomarcadores para quantificar aspectos do risco não genético, deverá ser possível obter uma previsão cada vez mais precisa do risco de doença individual e utilizar esta informação para triagem personalizada, prevenção e tratamento. O sucesso dependerá do desenvolvimento de modelos de risco que integrem de forma robusta estes diversos tipos de dados e da otimização das estratégias implementadas para garantir uma implementação eficaz.

A ausência de diretrizes baseadas em evidências para apoiar as recomendações de saúde continua a dificultar as aplicações clínicas dos dados genéticos. Em alguns países, esta situação é agravada pela confusão sobre o reembolso e pelas disparidades nos testes em toda a sociedade [148]. Muitos profissionais de saúde não têm experiência em medicina genómica e necessitam de educação e orientação para praticar no espaço em rápida evolução dos testes genéticos e genómicos [149]. Uma consequência destas dificuldades tem sido um mercado em expansão de testes diretos ao consumidor, controlado de forma variável por regulamentações específicas de cada país [150], que está a evoluir para além do foco na ascendência e nas características pessoais, rumo a modelos em que os indivíduos têm acesso direto a pedidos de médicos e conselheiros genéticos [151]. O risco de influência comercial neste modelo continua elevado. Existem preocupações sobre as consequências da divulgação irrestrita de dados genéticos de relevância clínica duvidosa ou inflacionada, e da infra-estrutura limitada para incluir estes resultados nos sistemas médicos convencionais.

Estes avanços fomentaram o debate sobre o valor da genética para o rastreio populacional, tanto para doenças monogénicas como para doenças complexas. É mais provável que o rastreio populacional de doenças monogénicas seja iniciado para condições para as quais as estimativas de risco são bem compreendidas e existem intervenções acionáveis (por exemplo, síndrome de Lynch e hipercolesterolemia familiar). A expansão para outras doenças requer uma melhor compreensão da penetrância dos alelos patogénicos em populações não seleccionadas [152] e cautela antes de alargar o rastreio a listas mais longas de genes que estão implicados de

forma menos segura na causa da doença [153]. À medida que certos países consideram a captura universal de dados genéticos de todo o genoma no nascimento ou mais tarde na vida, as principais questões dizem respeito às estratégias para divulgar esta informação aos cidadãos e às suas equipas médicas para apoiar os cuidados de saúde individuais.

Em última análise, as barreiras à medicina genómica são superadas mais diretamente através da demonstração da utilidade clínica na gestão de doenças e na tomada de decisões terapêuticas, com evidências de melhores resultados para os pacientes. Os cânceres hereditários fornecem vários exemplos, como o uso de testes *BRCA1/BRCA2* para informar o tratamento com inibidores de PARP em pacientes com câncer [154]. Há uma lista crescente de doenças para as quais um diagnóstico molecular resulta em intervenções específicas destinadas a melhorar os resultados dos pacientes (https://www.ncbi.nlm.nih.gov/books/NBK1116/) (alguns exemplos estão listados na Tabela 1) , e existem atualmente mais de 50 medicamentos aprovados pela FDA para doenças genéticas [155]. Embora a terapia genética tenha evoluído lentamente desde a sua introdução inicial, os avanços recentes na edição genética estão revigorando abordagens para tratar doenças através da manipulação dos defeitos genéticos subjacentes [156].

Tabela 1 Exemplos que mostram progresso na utilidade clínica dos testes genéticos

Esperando ansiosamente

Durante a próxima década, o desafio será optimizar e implementar em escala estratégias que utilizem a genética humana para aprofundar a compreensão da saúde e da doença e maximizar o benefício clínico dessas descobertas. A concretização destes objetivos exigirá o esforço concertado dos investigadores da academia e da indústria para provocar mudanças transformacionais numa série de domínios altamente interligados, por exemplo, através dos auspícios da recentemente criada Aliança Internacional para as Doenças Comuns (https://www.icda. biografia). Tais esforços serão direcionados para o estabelecimento de: (a) inventários abrangentes das relações genótipo-fenótipo entre populações e ambientes; (b) ensaios

sistemáticos da função em nível de variante e gene em todos os tipos de células, estados e exposições; (c) estratégias melhoradas e escaláveis para transformar este conhecimento básico em modelos moleculares, celulares e fisiológicos totalmente desenvolvidos de patogénese de doenças; e (d) aplicação desses conhecimentos biológicos para impulsionar novas opções preventivas e terapêuticas.

A primeira delas envolverá a documentação de todo o espectro da variação genética natural em todas as populações humanas, incluindo a captura de variantes estruturais e mutações somáticas que se acumulam com o envelhecimento [157,158], e a associação dessas variações com a cada vez mais rica variedade intermediária e relacionada à doença. características clínicas disponíveis através de biobancos e registros eletrônicos de saúde. Será particularmente importante incluir populações historicamente sub-representadas na investigação genómica, na sequência do trabalho pioneiro do consórcio H3Africa [159]. Como, com o tempo, os genomas sequenciados clinicamente ultrapassarão o número dos recolhidos no meio académico, as comunidades de investigação e de saúde terão de desenvolver uma abordagem harmonizada à genómica para transcender as fronteiras históricas. O progresso dependerá criticamente de plataformas e governação que reduzam as barreiras à integração de dados genéticos e fenotípicos entre estudos e países, juntamente com normas técnicas que sejam fiáveis, seguras e compatíveis com o panorama regulamentar internacional [160].

A interpretação mecanística das associações genéticas, particularmente aquelas em regiões reguladoras, será impulsionada pela anotação sistemática de variantes de sequências e genes para impacto funcional em tipos de células relevantes para a doença, permitindo o mapeamento de processos que contribuem para o desenvolvimento da doença no que diz respeito ao local (tecido e célula). tipo), tempo (fase de desenvolvimento) e contexto (influências externas) [161]. A aceleração dos esforços para caracterizar a composição celular dos tecidos através de ensaios unicelulares [115] aumentará a granularidade destas observações. Estudos de perturbação em larga escala em diversos modelos celulares e animais, juntamente com análises de variantes de codificação em humanos [53,54], proporcionarão confiança na inferência causal. Análises proteômicas e metabolômicas em larga

escala (em tecidos e fluidos biológicos) fornecerão uma ponte para vias a jusante [79, 80]. O acesso da investigação a esses dados funcionais, gerados em escala, deverá reduzir as barreiras à inferência mecanicista, fornecer um contexto de todo o sistema e permitir que os investigadores concentrem a validação de laboratório húmido nas experiências mais críticas. Coletivamente, estes esforços apoiarão a compilação de um catálogo sistemático de redes e processos-chave que influenciam a fisiologia normal e o desenvolvimento de doenças e informarão uma taxonomia molecular revista das doenças.

Este conhecimento reforçará a contribuição essencial da genética humana para a identificação e priorização de alvos para o desenvolvimento terapêutico [89, 162]. Os insights sobre a eficácia da perturbação do alvo e o potencial para eventos adversos, aliados à caracterização de biomarcadores traduzíveis, fornecem formas de aumentar a eficiência dos canais de desenvolvimento de medicamentos [162]. Dada a importância clínica de retardar a progressão da doença [163], os esforços de descoberta de alvos necessitarão cada vez mais de abranger a genética da progressão da doença e da resposta ao tratamento, uma vez que estes podem envolver processos distintos daqueles capturados pelos estudos do início da doença.

Paralelamente, a utilização clínica da genética humana beneficiará do progresso no sentido da determinação universal de sequências genómicas individuais construídas através de uma combinação de expansão de biobancos e acesso direto nos sistemas de saúde. Isto impulsionará aplicações clínicas que vão além do foco atual no sequenciamento neonatal, no diagnóstico mendeliano e no sequenciamento de tumores somáticos [164]. Em particular, as melhorias na derivação de pontuações poligénicas aumentarão a previsão de risco para características multifatoriais, fornecerão uma base molecular para a classificação de doenças, apoiarão a descoberta de biomarcadores e a otimização terapêutica e contribuirão para a compreensão da penetrância variável de condições monogénicas [69]. A implementação da medicina genómica como uma componente de rotina dos cuidados clínicos em diversos ambientes de saúde exigirá inevitavelmente investimento na formação de profissionais de saúde e atenção às estratégias ideais para devolver os resultados genéticos aos pacientes.

A herdabilidade limitada de muitas características multifatoriais restringe a precisão clínica disponível apenas a partir dos dados genéticos. Isto impulsionará esforços para integrar informações sobre o ambiente pessoal, estilo de vida e comportamento, e para combinar informações prognósticas e preditivas sobre o risco de doenças com medidas longitudinais do estado molecular e clínico que acompanham a jornada de um indivíduo da saúde à doença. A genética humana também apoiará, dado o seu potencial único de inferência causal, a identificação dos factores de risco não genéticos (muitas vezes modificáveis) que contribuem directamente para a predisposição e o desenvolvimento de doenças [165]. À medida que o desempenho da pontuação poligénica melhora, a análise de indivíduos que apresentam divergência acentuada entre a predisposição genética e os resultados clínicos do mundo real deve definir as exposições (tais como escolhas de estilo de vida ou microbioma intestinal) cuja contribuição para a causa da doença permanece obscura [166].

Coletivamente, espera-se que estes desenvolvimentos acelerem a personalização da prestação de cuidados de saúde. Desde que os custos sejam sustentáveis, poderá surgir uma perspetiva mais preventiva da saúde, gerida através de uma vigilância proativa genómica, clínica e do estilo de vida, utilizando pontuações de risco, biomarcadores complexos, biópsias líquidas e dispositivos vestíveis. Uma melhor compreensão da heterogeneidade etiológica, dos padrões de partilha do risco genético entre as doenças, da variação na resposta terapêutica e do risco de acontecimentos adversos melhorará o direcionamento das intervenções preventivas e terapêuticas [167]. A nível da população, as estratégias de intervenção procurarão combinar estratégias que abrangem toda a população e estratégias específicas para obter os melhores resultados [168]. Será fundamental garantir que estes benefícios estejam disponíveis para o maior número possível de pessoas, para que a genómica reduza, em vez de agravar, as disparidades de saúde nacionais e globais [55,169] (Caixa 1).

Os desenvolvimentos descritos acima representam variações do tema da “leitura” do genoma. A capacidade emergente de bloquear esta leitura (por exemplo, através de terapias com siRNA [170]) ou mesmo de “escrever” o genoma (através da edição

CRISPR) promete ser igualmente transformadora, proporcionando novas oportunidades para corrigir, e até curar, a doença mendeliana. Avanços espetaculares no desenvolvimento de novas estratégias terapêuticas são prováveis para muitas doenças, com base, por exemplo, na manipulação celular ex vivo [171] ou na edição de células somáticas in vivo [172] .

É importante ressaltar que os desenvolvimentos na medicina genómica precisam de prosseguir num quadro bioético para investigação e utilização clínica que reconheça a relevância pessoal da genética humana e a importância crítica do consentimento autónomo e da protecção da privacidade, minimizando ao mesmo tempo as consequências adversas do excepcionalismo genético. A governação precisa de reafirmar os direitos dos cidadãos de fazerem contribuições individuais para o progresso científico através da participação na investigação e encorajar o intercâmbio responsável de dados para fins clínicos e de investigação.

Mostre mais

Perspectivas futuras

Nas últimas duas décadas, a compreensão da base genética das doenças humanas foi transformada por uma combinação de avanços tecnológicos e analíticos espectaculares, pelo compromisso colaborativo com o desenvolvimento de recursos fundamentais e pela recolha e análise de grandes quantidades de dados genéticos, moleculares e clínicos. . Os conhecimentos biológicos derivados destes dados são, cada vez mais, impulsionadores da inovação translacional, e o alargamento do acesso pessoal a dados genéticos e moleculares em grande escala promete remodelar os cuidados médicos.

No entanto, para que todo o potencial da medicina genómica seja concretizado, será necessário um esforço colaborativo sustentado em várias frentes para garantir que a capacidade de gerar mapas cada vez mais detalhados das relações entre a variação da

sequência e os fenótipos biomédicos proporciona uma compreensão abrangente da doença. mecanismos que podem ser traduzidos nos medicamentos de amanhã.

Referências

1. Consórcio Internacional de Sequenciamento do Genoma Humano. Sequenciamento inicial e análise do genoma humano. *Natureza* **409** , 860–921 (2001). **Este artigo descreve as primeiras análises do rascunho da sequência do genoma humano montado ao longo da década anterior: lançou a genética humana moderna e representa um tributo ao poder da ciência colaborativa.**

2. Consórcio Internacional HapMap. O Projeto Internacional HapMap. *Natureza* **426** , 789–796 (2003). **O Consórcio HapMap desenvolveu os primeiros mapas genômicos de variação de sequência comum, usando esta informação para traçar a estrutura haplotípica desta variação em três grandes agrupamentos ancestrais (da Europa, Leste Asiático e África).**

3. Consórcio do Projeto 1000 Genomas. Uma referência global para a variação genética humana. *Natureza* **526** , 68–74 (2015).

4. O Consórcio de Referência de Haplótipos. Um painel de referência de 64.976 haplótipos para imputação de genótipos. *Nat. Geneta* . **48** , 1279–1283 (2016).

5. Lek, M. et al. Análise da variação genética codificadora de proteínas em 60.706 humanos. *Natureza* **536** , 285–291 (2016).

6. Karczewski, KJ et al. A variação em 141.456 exomas e genomas humanos revela o espectro de intolerância à perda de função em genes codificadores de proteínas humanas. Pré-impressão em *bioRxiv* https://doi.org/10.1101/531210 (2019). **O relatório mais recente do projeto de base de dados de agregação de genoma (gnomAD) que combina dados de sequências humanas numa escala sem**

precedentes para caracterizar variantes raras e de alto impacto e a sua relação com a saúde e a doença.

7. Kremer, B. et al. Um estudo mundial da mutação da doença de Huntington. A sensibilidade e especificidade da medição de repetições CAG. *N. Engl. J. Med* . **330** , 1401–1406 (1994).

8. Collins, FS Identificando genes de doenças humanas por clonagem posicional. *Palestra Harvey* . **86** , 149–164 (1990–1991).

9. Gusella, JF e MacDonald, ME Doença de Huntington e trinucleotídeos de repetição. *N. Engl. J. Med* . **330** , 1450–1451 (1994).

10. Vissers, LE et al. Hibridização genômica comparativa baseada em array para a detecção genômica de anormalidades cromossômicas submicroscópicas. *Sou. J. Hum. Geneta* . **73** , 1261–1270 (2003).

11. Ng, SB et al. Captura direcionada e sequenciamento massivamente paralelo de 12 exomas humanos. *Natureza* **461** , 272–276 (2009).

12. Wheeler, DA et al. O genoma completo de um indivíduo por sequenciamento massivamente paralelo de DNA. *Natureza* **452** , 872–876 (2008).

13. Vissers, LE et al. Um novo paradigma para o retardo mental. *Nat. Geneta* . **42** , 1109–1112 (2010).

14. Decifrando o Estudo dos Transtornos do Desenvolvimento. Prevalência e arquitetura de mutações de novo em distúrbios do desenvolvimento. *Natureza*

542 , 433–438 (2017).

15. Splinter, K. et al. Efeito do diagnóstico genético em pacientes com doença previamente não diagnosticada. *N. Engl. J. Med* . **379** , 2131–2139 (2018).

16. Decifrando o Estudo dos Transtornos do Desenvolvimento. Descoberta em larga escala de novas causas genéticas de distúrbios do desenvolvimento. *Natureza* **519** , 223–228 (2015).

17. Firth, HV et al. DECIPHER: banco de dados de desequilíbrio cromossômico e fenótipo em humanos usando recursos de conjunto. *Sou. J. Hum. Geneta* . **84** , 524–533 (2009).

18. Philippakis, AA et al. The Matchmaker Exchange: uma plataforma para descoberta de genes de doenças raras. *Zumbir. Mutação.* **36** , 915–921 (2015).

19. Boicote, KM et al. Cooperação internacional para permitir o diagnóstico de todas as doenças genéticas raras. *Sou. J. Hum. Geneta* . **100** , 695–705 (2017).

20. Kennedy, J. et al. Síndrome KAT6A: correlação genótipo-fenótipo em 76 pacientes com variantes patogênicas *do KAT6A* . *Geneta. Med* . **21** , 850–860 (2019).

21. Amberger, JS, Bocchini, CA, Schiettecatte, F., Scott, AF & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM), um catálogo online de genes humanos e doenças genéticas. *Ácidos Nucleicos Res* . **43** , D789–D798 (2015).

22. Landrum, MJ et al. ClinVar: arquivo público de relações entre variação de sequência e fenótipo humano. *Ácidos Nucleicos Res* . **42** , D980–D985 (2014).

23. Rehm, HL et al. ClinGen – o recurso do genoma clínico. *N. Engl. J. Med* . **372** , 2235–2242 (2015).

24. Minikel, EV et al. Quantificando a penetrância da doença do príon usando grandes coortes de controle populacional. *Ciência. Trad. Med* . **8** , 322ra9 (2016).

25. Flannick, J. et al. Avaliando os efeitos fenotípicos na população geral de variantes raras em genes para uma forma mendeliana dominante de diabetes. *Nat. Geneta* . **45** , 1380–1385 (2013).

26. Bečanović, K. et al. Um SNP no promotor *HTT* altera a ligação do NF-κB e é um modificador genético bidirecional da doença de Huntington. *Nat. Neurosci* . **18** , 807–816 (2015).

27. Castel, SE et al. A penetrância modificada de variantes de codificação pela variação *regulatória* cis contribui para o risco de doenças. *Nat. Geneta* . **50** , 1327–1334 (2018).

28. MacArthur, DG et al. Diretrizes para investigar a causalidade de variantes de sequência em doenças humanas. *Natureza* **508** , 469–476 (2014).

29. Findlay, GM et al. Classificação precisa de variantes de *BRCA1* com edição de saturação do genoma. *Natureza* **562** , 217–222 (2018). **Este artigo descreve um esforço para obter avaliações funcionais massivamente paralelas de todas as possíveis variantes de sequência em um gene causal para o câncer de mama**

familiar e, assim, fornecer previsões mais confiáveis de significância clínica para aqueles que carregam uma mutação anteriormente não observada.

30. Cummings, BB et al. Melhorando o diagnóstico genético na doença mendeliana com sequenciamento do transcriptoma. *Ciência. Trad. Med* . **9** , eaal5209 (2017).

31. Butcher, DT et al. Síndromes CHARGE e Kabuki: assinaturas de metilação do DNA específicas do gene identificam mecanismos epigenéticos que ligam essas condições clinicamente sobrepostas. *Sou. J. Hum. Geneta* . **100** , 773–788 (2017).

32. Resumindo, PJ et al. Mutações de novo em elementos reguladores em distúrbios do neurodesenvolvimento. *Natureza* **555** , 611–616 (2018).

33. Jaganathan, K. et al. Previsão de emenda da sequência primária com aprendizado profundo. *Célula* **176** , 535–548 (2019).

34. Hugot, JP et al. Associação de variantes repetidas ricas em leucina *NOD2* com suscetibilidade à doença de Crohn. *Natureza* **411** , 599–603 (2001).

35. Ogura, Y. et al. Uma mutação frameshift no *NOD2* associada à suscetibilidade à doença de Crohn. *Natureza* **411** , 603–606 (2001).

36. Barbosa, J., Chern, MM, Noreen, H. & Anderson, VE Análise da ligação entre o sistema principal de histocompatibilidade e diabetes juvenil dependente de insulina em famílias multiplex. Reanálise de dados. *J. Clin. Invista* . **62** , 492–495 (1978).

37. Risch, N. & Merikangas, K. O futuro dos estudos genéticos de doenças humanas complexas. *Ciência* **273** , 1516–1517 (1996). **Um comentário altamente influente que mudou grande parte do campo da genética de características complexas de ligação para associação e imaginou GWAS vários anos antes de a tecnologia torná-los uma realidade.**

38. Klein, RJ et al. Polimorfismo do fator H do complemento na degeneração macular relacionada à idade. *Ciência* **308** , 385–389 (2005). **Indiscutivelmente o primeiro GWAS completo, este estudo demonstrou o potencial de pesquisas genômicas agnósticas para destacar uma biologia inteiramente nova – neste caso, um papel do complemento na patogênese da degeneração macular.**

39. Duerr, RH et al. Um estudo de associação genômica identifica *IL23R* como um gene de doença inflamatória intestinal. *Ciência* **314** , 1461–1463 (2006).

40. Consórcio de controle de caso Wellcome Trust. Estudo de associação de todo o genoma de 14.000 casos de sete doenças comuns e 3.000 controles compartilhados. *Natureza* **447** , 661–678 (2007). **Este artigo, que descreve o que foi, na época, o maior GWAS já realizado, demonstrou a ampla aplicabilidade da abordagem e preparou o cenário para a década de descoberta de GWAS que se seguiu.**

41. Mahajan, A. et al. Mapeamento preciso de loci de diabetes tipo 2 para resolução de variante única usando imputação de alta densidade e mapas de epigenoma específicos de ilhotas. *Nat. Geneta* . **50** , 1505–1513 (2018).

42. Astle, WJ et al. A paisagem alélica da variação das características das células sanguíneas humanas e suas ligações com doenças complexas comuns. *Célula* **167** , 1415–1429 (2016).

43. Lango Allen, H. et al. Centenas de variantes agrupadas em loci genômicos e vias biológicas afetam a altura humana. *Natureza* **467** , 832–838 (2010).

44. Buniello, A. et al. O catálogo NHGRI – EBI GWAS de estudos de associação de todo o genoma publicados, matrizes direcionadas e estatísticas resumidas 2019. *Nucleic Acids Res* . **47** , D1005–D1012 (2019).

45. Fuchsberger, C. et al. A arquitetura genética do diabetes tipo 2. *Natureza* **536** , 41–47 (2016).

46. Consórcio UK10K. O projeto UK10K identifica variantes raras na saúde e na doença. *Natureza* **526** , 82–90 (2015).

47. Manólio, TA et al. Encontrando a falta de herdabilidade de doenças complexas. *Natureza* **461** , 747–753 (2009).

48. Wainschtein, P. et al. Recuperação da herdabilidade das características a partir de dados de sequência completa do genoma. Pré-impressão em *bioRxiv* https://doi.org/10.1101/588020 (2019).

49. Flannick, J. et al. Mutações de perda de função em *SLC30A8* protegem contra diabetes tipo 2. *Nat. Geneta* . **46** , 357–363 (2014).

50. Moltke, I. et al. Uma variante comum *do TBC1D4* da Groenlândia confere resistência muscular à insulina e diabetes tipo 2. *Natureza* **512** , 190–193 (2014).

51. Nioi, P. et al. Variante *ASGR1* associada a um risco reduzido de doença arterial coronariana. *N. Engl. J. Med* . **374** , 2131–2141 (2016).

52. Pollin, TI et al. Uma mutação nula no *APOC3* humano confere um perfil lipídico plasmático favorável e aparente cardioproteção. *Ciência* **322** , 1702–1705 (2008).

53. Narasimhan, VM et al. Efeitos na saúde e na população de nocautes genéticos raros em humanos adultos com pais aparentados. *Ciência* **352** , 474–477 (2016). **Uma ilustração poderosa dos insights que podem ser obtidos com o estudo de indivíduos que são portadores homozigotos de alelos nulos em genes de interesse biomédico.**

54. Saleheen, D. et al. Nocautes humanos e análise fenotípica em uma coorte com alto índice de consanguinidade. *Natureza* **544** , 235–239 (2017).

55. Martin, AR et al. O uso clínico dos atuais escores de risco poligênico pode exacerbar as disparidades de saúde. *Nat. Geneta* . **51** , 584–591 (2019). **Uma destilação elegante das maneiras pelas quais o foco histórico da descoberta genética entre populações de ascendência europeia tem consequências adversas tanto para a descoberta como para a tradução.**

56. Zeng, J. et al. Assinaturas de seleção negativa na arquitetura genética de características complexas humanas. *Nat. Geneta* . **50** , 746–753 (2018).

57. Genovese, G. et al. Associação de variantes tripanolíticas do ApoL1 com doença renal em afro-americanos. *Ciência* **329** , 841–845 (2010).

58. Luzzatto, L., Usanga, FA & Reddy, S. Glóbulos vermelhos deficientes em glicose-6-fosfato desidrogenase: resistência à infecção por parasitas da malária. *Ciência* **164** , 839–842 (1969).

59. Wojcik, GL et al. Análises genéticas de diversas populações melhoram a descoberta de características complexas. *Natureza* **570** , 514–518 (2019).

60. Farh, KK et ai. Mapeamento fino genético e epigenético de variantes causais de doenças autoimunes. *Natureza* **518** , 337–343 (2015). **Uma demonstração inicial do valor da sobreposição de dados de associação de todo o genoma com mapas regulatórios específicos de tecidos para definir os tipos de células e tecidos que provavelmente estão impulsionando a patologia da doença.**

61. Freund, MK et al. Enriquecimento específico do fenótipo de genes do distúrbio mendeliano próximos às regiões GWAS em 62 características complexas. *Sou. J. Hum. Geneta* . **103** , 535–552 (2018).

62. Boyle, EA, Li, YI & Pritchard, JK Uma visão ampliada de características complexas: de poligênico a omnigênico. *Célula* **169** , 1177–1186 (2017).

63. de Lange, KM et al. O estudo de associação de todo o genoma implica a ativação imunológica de múltiplos genes da integrina na doença inflamatória intestinal. *Nat. Geneta* . **49** , 256–261 (2017).

64. Sekar, A. et al. Risco de esquizofrenia devido à variação complexa do componente do complemento 4. *Nature* **530** , 177–183 (2016).

65. Jostins, L. et al. As interações hospedeiro-micróbio moldaram a arquitetura genética da doença inflamatória intestinal. *Natureza* **491** , 119–124 (2012).

66. Richardson, TG, Harrison, S., Hemani, G. & Davey Smith, G. Um atlas de associações de pontuação de risco poligênico para destacar possíveis relações causais em todo o fenômeno humano. *eLife* **8** , e43657 (2019).

67. Dendrou, CA et al. Resolvendo diferenças de genótipo para fenótipo do locus TYK2 na autoimunidade. *Ciência. Trad. Med* . **8** , 363ra149 (2016).

68. Voight, BF et al. Colesterol HDL plasmático e risco de infarto do miocárdio: um estudo mendeliano de randomização. *Lancet* **380** , 572–580 (2012).

69. Khera, AV et al. Pontuações poligênicas em todo o genoma para doenças comuns identificam indivíduos com risco equivalente a mutações monogênicas. *Nat. Geneta* . **50** , 1219–1224 (2018). **Este artigo utilizou dados genéticos em larga escala para demonstrar o potencial clínico dos escores poligênicos que podem ser construídos para muitas doenças comuns, enfatizando que, em algumas situações, o risco de doença ao longo da vida para aqueles com os escores mais altos se aproxima do risco de doença monogênica estabelecida. .**

70. Mahmood, SS, Levy, D., Vasan, RS & Wang, TJ O Framingham Heart Study e a epidemiologia das doenças cardiovasculares: uma perspectiva histórica. *Lancet* **383** , 999–1008 (2014).

71. Gudbjartsson, DF et al. Variantes que conferem risco de fibrilação atrial no cromossomo 4q25. *Natureza* **448** , 353–357 (2007).

72. Thorgeirsson, TE et al. Uma variante associada à dependência da nicotina, câncer de pulmão e doença arterial periférica. *Natureza* **452** , 638–642 (2008).

73. Collins, FS & Varmus, H. Uma nova iniciativa em medicina de precisão. *N. Engl. J. Med* . **372** , 793–795 (2015).

74. Bycroft, C. et al. O recurso UK Biobank com fenotipagem profunda e dados genômicos. *Natureza* **562** , 203–209 (2018).

75. Nagai, A. et al. Visão geral do Projeto BioBank Japan: desenho e perfil do estudo. *J. Epidemiol* . **27** , S2–S8 (2017).

76. Chen, Z. et al. China Kadoorie Biobank de 0,5 milhão de pessoas: métodos de pesquisa, características iniciais e acompanhamento de longo prazo. *Internacional J. Epidemiol* . **40** , 1652–1666 (2011).

77. Kohane, IS Usando registros eletrônicos de saúde para impulsionar a descoberta da genômica de doenças. *Nat. Rev.* **12** , 417–428 (2011).

78. Abul-Husn, NS & Kenny, EE Medicina personalizada e o poder dos registros eletrônicos de saúde. *Célula* **177** , 58–69 (2019).

79. Sun, BB et al. Atlas genômico do proteoma plasmático humano. *Natureza* **558** , 73–79 (2018).

80. Shin, SY et al. Um atlas de influências genéticas nos metabólitos do sangue humano. *Nat. Geneta* . **46** , 543–550 (2014).

81. Sung, YJ et al. Um estudo de genoma multi-ancestral incorporando interações genéticas de fumo identifica vários novos loci para pressão de pulso e pressão arterial média. *Zumbir. Mol. Geneta* . **28** , ddz070 (2019).

82. Pendergrass, SA e Crawford, DC Uso de registros eletrônicos de saúde para gerar fenótipos para pesquisa. *Curr. Protocolo. Zumbir. Geneta* . **100** , e80 (2019).

83. Denny, JC et al. PheWAS: demonstrando a viabilidade de uma varredura ampla do fenômeno para descobrir associações entre genes e doenças. *Bioinformática* **26** , 1205–1210 (2010).

84. Liu, Z. et al. A análise de randomização mendeliana disseca a relação entre DHGNA, DM2 e obesidade e fornece implicações para a medicina de precisão. Pré-impressão em *bioRxiv* https://doi.org/10.1101/657734 (2019).

85. Cotsapas, C. et al. Compartilhamento generalizado de efeitos genéticos em doenças autoimunes. *PLoS Genet* . **7** , e1002254 (2011).

86. Kanai, M. et al. A análise genética de características quantitativas na população japonesa liga os tipos de células a doenças humanas complexas. *Nat. Geneta* . **50** , 390–400 (2018).

87. Gudmundsson, J. et al. Duas variantes no cromossomo 17 conferem risco de câncer de próstata, e a do *TCF2* protege contra o diabetes tipo 2. *Nat. Geneta* . **39** , 977–983 (2007).

88. Abul-Husn, NS et al. Uma variante *HSD17B13* truncada de proteína e proteção contra doença hepática crônica. *N. Engl. J. Med* . **378** , 1096–1106 (2018).

89. Plenge, RM, Scolnick, EM & Altshuler, D. Validando alvos terapêuticos através da genética humana. *Nat. Rev. Drug Discov* . **12** , 581–594 (2013).

90. Nielsen, JB et al. A descoberta genômica conduzida por biobancos produz novos insights sobre a biologia da fibrilação atrial. *Nat. Geneta* . **50** , 1234–1239 (2018).

91. Flannick, J. et al. Sequenciamento do exoma de 20.791 casos de diabetes tipo 2 e 24.440 controles. *Natureza* **570** , 71–76 (2019).

92. Van Hout, CV et al. Sequenciamento completo do exoma e caracterização da variação de codificação em 49.960 indivíduos no Biobanco do Reino Unido. Pré-impressão em *bioRxiv* https://doi.org/10.1101/572347 (2019).

93. Khera, AV et al. Sequenciamento do genoma completo para caracterizar contribuições monogênicas e poligênicas em pacientes hospitalizados com infarto do miocárdio de início precoce. *Circulação* **139** , 1593–1602 (2019).

94. Grarup, N. et al. Variantes de perda de função no *ADCY3* aumentam o risco de obesidade e diabetes tipo 2. *Nat. Geneta* . **50** , 172–174 (2018).

95. Cohen, JC, Boerwinkle, E., Mosley, TH Jr & Hobbs, HH Variações de sequência em *PCSK9* , LDL baixo e proteção contra doenças coronarianas. *N. Engl. J. Med* . **354** , 1264–1272 (2006).

96. Rivas, MA et al. O resequenciamento profundo dos loci GWAS identifica variantes raras independentes associadas à doença inflamatória intestinal. *Nat. Geneta* . **43** , 1066–1073 (2011).

97. Chong, JX et al. A base genética dos fenótipos mendelianos: descobertas, desafios e oportunidades. *Sou. J. Hum. Geneta* . **97** , 199–215 (2015).

98. Goldstein, JL & Brown, MS Um século de colesterol e doenças coronárias: de placas a genes e estatinas. *Célula* **161** , 161–172 (2015).

99. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lições do estudo de doenças monogênicas para doenças comuns. *Zumbir. Mol. Geneta* . **15** , R67–R74 (2006).

100. Bulik-Sullivan, B. et al. Um atlas de correlações genéticas entre doenças e características humanas. *Nat. Geneta* . **47** , 1236–1241 (2015).

101. Musunuru, K. et al. Da variante não codificante ao fenótipo via *SORT1* no locus do colesterol 1p13. *Natureza* **466** , 714–719 (2010). **Amplamente aceito como o primeiro artigo a descrever a dissecação mecanicista detalhada de um locus de característica complexo e ainda um dos relativamente poucos exemplos de sucesso neste empreendimento.**

102. Claussnitzer, M. et al. Circuito variante da obesidade FTO e escurecimento dos adipócitos em humanos. *N. Engl. J. Med* . **373** , 895–907 (2015).

103. Smemo, S. et al. Variantes associadas à obesidade dentro do *FTO* formam conexões funcionais de longo alcance com *o IRX3* . *Natureza* **507** , 371–375 (2014).

104. Gupta, RM et al. Uma variante genética associada a cinco doenças vasculares é um regulador distal da expressão do gene da endotelina-1. *Célula* **170** , 522–533 (2017).

105. Pequeno, KS et al. Variantes regulatórias no *KLF14* influenciam o risco de diabetes tipo 2 através de um efeito específico da mulher no tamanho dos adipócitos e na composição corporal. *Nat. Geneta* . **50** , 572–580 (2018).

106. Lin, YT et al. *APOE4* causa alterações moleculares e celulares generalizadas associadas aos fenótipos da doença de Alzheimer em tipos de células cerebrais humanas derivadas de iPSC. *Neurônio* **98** , 1294 (2018).

107. Consórcio do Projeto ENCODE. Uma enciclopédia integrada de elementos de DNA no genoma humano. *Natureza* **489** , 57–74 (2012).

108. Roteiro Epigenomics Consortium et al. Análise integrativa de 111 epigenomas humanos de referência. *Natureza* **518** , 317–330 (2015). **O principal artigo do NIH Roadmap Epigenomics Consortium, descrevendo uma análise detalhada dos perfis epigenômicos de mais de 100 tipos de células humanas e caracterizando como esses padrões regulatórios se relacionam com a regulação genética, a diferenciação celular e as doenças humanas.**

109. Tansey, KE, Cameron, D. & Hill, MJ O risco genético para a doença de Alzheimer está concentrado em macrófagos específicos e redes transcricionais microgliais. *Genoma Med* . **10** , 14 (2018).

110. Gamazon, ER et al. Usando um atlas de regulação genética em 44 tecidos humanos para informar variações complexas associadas a doenças e características. *Nat. Geneta* . **50** , 956–967 (2018).

111. Ongen, H. et al. Estimando os tecidos causais para características e doenças complexas. *Nat. Geneta* . **49** , 1676–1683 (2017).

112. Consórcio GTEx. Efeitos genéticos na expressão gênica em tecidos humanos. *Natureza* **550** , 204–213 (2017). **Um dos vários artigos que descrevem os resultados obtidos no projeto Genotype-Tissue Expression (GTEx), que utilizou dados de sequência de RNA de mais de 40 tecidos humanos em várias centenas de indivíduos para explorar a relação entre a variação da sequência de DNA e a expressão específica do tecido.**

113. Chiou, J. et al. A acessibilidade à cromatina de célula única revela programas regulatórios de risco de diabetes específicos do tipo e do estado das células das ilhotas pancreáticas. Pré-impressão em *bioRxiv* https://doi.org/10.1101/693671 (2019).

114. Consórcio HuBMAP. O corpo humano em resolução celular: o Programa NIH Human Biomolecular Atlas. *Natureza* **574** , 187–192 (2019).

115. Regev, A. et al. O Atlas da Célula Humana. *eLife* **6** , e27041 (2017).

116. Lloyd, KC Um recurso de rato nocaute para a comunidade de pesquisa biomédica. *Ana. NY Acad. Ciência* . **1245** , 24–26 (2011).

117. Finucane, HK et al. O enriquecimento de herdabilidade de genes especificamente expressos identifica tecidos e tipos de células relevantes para a doença. *Nat. Geneta* . **50** , 621–629 (2018).

118. Delaneau, O. et al. As interações tridimensionais da cromatina medeiam os efeitos genéticos na expressão gênica. *Ciência* **364** , eaat8266 (2019).

119. Zhou, J. & Troyanskaya, OG Predição de efeitos de variantes não codificantes com modelo de sequência baseado em aprendizagem profunda. *Nat. Métodos* **12** , 931–934 (2015).

120. Kelley, DR, Snoek, J. & Rinn, JL Basset: aprendendo o código regulador do genoma acessível com redes neurais convolucionais profundas. *Res. do Genoma* . **26** , 990–999 (2016).

121. Patwardhan, RP et al. Análise de alta resolução de elementos reguladores de DNA por mutagênese de saturação sintética. *Nat. Biotecnologia* . **27** , 1173–1175 (2009).

122. Melnikov, A. et al. Dissecação sistemática e otimização de intensificadores indutíveis em células humanas usando um ensaio repórter massivamente paralelo. *Nat. Biotecnologia* . **30** , 271–277 (2012).

123. Wang, X. et al. Dissecção funcional de alta resolução em todo o genoma de regiões reguladoras da transcrição e nucleotídeos em humanos. *Nat. Comum* . **9** , 5380 (2018).

124. Potting, C. et al. A triagem CRISPR de todo o genoma para reguladores PARKIN revela a repressão transcricional como um determinante da mitofagia. *Processo. Acad. Nacional. Ciência. EUA* **115** , E180–E189 (2018).

125. Gasperini, M. et al. Uma estrutura genômica ampla para mapear a regulação genética por meio de telas genéticas celulares. *Célula* **176** , 377–390 (2019).

126. Fulco, CP et al. Modelo atividade por contato de regulação intensificador-promotor de milhares de perturbações CRISPR. *Nat. Geneta* . **51** , 1664–1669 (2019).

127. Komor, AC, Kim, YB, Packer, MS, Zuris, JA e Liu, DR Edição programável de uma base alvo em DNA genômico sem clivagem de DNA de fita dupla. *Natureza* **533** , 420–424 (2016).

128. Feldman, D. et al. Telas ópticas agrupadas em células humanas. *Célula* **179** , 787–799 (2019).

129. Dixit, A. et al. Perturb-Seq: dissecando circuitos moleculares com perfil escalonável de RNA unicelular de telas genéticas agrupadas. *Célula* **167** , 1853–1866 (2016).

130. Datlinger, P. et al. Triagem CRISPR agrupada com leitura de transcriptoma unicelular. *Nat. Métodos* **14** , 297–301 (2017).

131. Ding, L. et al. Perspectiva sobre processos oncogênicos no final do início da genômica do câncer. *Célula* **173** , 305–320 (2018).

132. Chae, YK et al. Caminho em direção à oncologia de precisão: revisão de estudos de terapia direcionada e ferramentas para auxiliar na definição da "acionabilidade" de uma lesão molecular e suporte ao manejo do paciente. *Mol. Câncer Ter* . **16** , 2645–2655 (2017).

133. Shearer, AE & Smith, RJ Sequenciamento massivamente paralelo para diagnóstico genético de perda auditiva: o novo padrão de atendimento. *Otolaringol. Cirurgia de cabeça e pescoço* . **153** , 175–182 (2015).

134. Yang, Y. et al. Achados moleculares entre pacientes encaminhados para sequenciamento clínico do exoma completo. *Geléia. Med. Assoc* . **312** , 1870–1879 (2014).

135. Friedman, JM et al. Sequenciamento de todo o genoma em bebês com doenças agudas: aplicação crítica da medicina genômica? *Geneta. Med* . **21** , 498–504 (2019).

136. Clark, MM et al. Diagnóstico de doenças genéticas em crianças gravemente doentes por meio de sequenciamento rápido do genoma completo e fenotipagem e interpretação automatizadas. *Ciência. Trad. Med* . **11** , eaat6177 (2019).

137. Eldomery, MK et al. Lições aprendidas com análises adicionais de pesquisas de casos de exoma clínico não resolvidos. *Genoma Med* . **9** , 26 (2017).

138. Boicote, KM et al. Um diagnóstico para todas as doenças genéticas raras: o horizonte e as próximas fronteiras. *Célula* **177** , 32–37 (2019).

139. Liu, S. et al. Análises genômicas de testes pré-natais não invasivos revelam associações genéticas, padrões de infecções virais e história da população chinesa. *Célula* **175** , 347–359 (2018).

140. Thung, DT, Beulen, L., Hehir-Kwa, J. & Faas, BH Implementação de sequenciamento massivamente paralelo do genoma inteiro para testes pré-natais não invasivos em laboratórios. *Perito Rev. Mol. Diagnóstico* . **15** , 111–124 (2015).

141. Lo, JO, Cori D, F., Norton, ME & Caughey, AB Testes pré-natais não invasivos. *Obsteto. Ginecol. Sobreviver* . **69** , 89–99 (2014).

142. Watson, MS et al. Triagem neonatal: rumo a um painel e sistema de triagem uniformes – resumo executivo. *Pediatria* **117** , S296–S307 (2006).

143. Currier, RJ et al. Sequenciamento genômico na triagem neonatal para fibrose cística: o que funciona melhor, painéis de mutação CFTR predefinidos de dois níveis ou painel CFTR de segundo nível seguido de sequenciamento de terceiro nível? *Geneta. Med* . **19** , 1159–1163 (2017).

144. Bauer, PE "Diga-lhes que não é tão ruim": triagem pré-natal para síndrome de Down e preconceito em relação ao aborto. *Intelecto. Dev. Desativar* . **46** , 247–251 (2008).

145. Volpi, S. et al. Direções de pesquisa na implementação clínica da farmacogenômica: uma visão geral dos programas e projetos dos EUA. *Clin.*

Farmacol. Lá . **103** , 778–786 (2018).

146. Consórcio Internacional de Esquizofrenia. A variação poligênica comum contribui para o risco de esquizofrenia e transtorno bipolar. *Natureza* **460** , 748–752 (2009).

147. Evans, DM, Visscher, PM & Wray, NR Aproveitar as informações contidas nos estudos de associação de todo o genoma para melhorar a previsão individual do risco de doenças complexas. *Zumbir. Mol. Geneta* . **18** , 3525–3531 (2009).

148. Kurian, AW et al. Testes genéticos e resultados em uma coorte populacional de pacientes com câncer de mama e pacientes com câncer de ovário. *J. Clin. Oncol* . **37** , 1305–1315 (2019).

149. Katz, SJ et al. Associação do cirurgião assistente com variação no recebimento de exames genéticos após diagnóstico de câncer de mama. *JAMA Surg* . **153** , 909–916 (2018).

150. Kalokairinou, L. et al. Legislação sobre testes genéticos diretos ao consumidor na Europa: um cenário regulatório fragmentado. *J. Comunidade Genet* . **9** , 117–132 (2018).

151. Allyse, MA, Robinson, DH, Ferber, MJ & Sharp, RR Teste Direto ao Consumidor 2.0: modelos emergentes de testes genéticos diretos ao consumidor. *Mayo Clin. Proc* . **93** , 113–120 (2018).

152. Wright, CF et al. Avaliação da patogenicidade, penetrância e expressividade de possíveis variantes causadoras de doenças em um ambiente populacional. *Sou. J. Hum. Geneta* . **104** , 275–286 (2019).

153. Conselho de Administração da ACMG. O uso de recomendações de resultados secundários da ACMG para triagem da população em geral: uma declaração de política do Colégio Americano de Genética Médica e Genômica (ACMG). *Geneta. Med* . **21** , 1467–1468 (2019).

154. George, A., Kaye, S. & Banerjee, S. Entrega de testes BRCA generalizados e inibição de PARP para pacientes com câncer de ovário. *Nat. Rev. Oncol* . **14** , 284–296 (2017).

155. CentroWatch. *Medicamentos aprovados pela FDA para doenças genéticas* (CenterWatch); https://www.centerwatch.com/drug-information/fda-approved-drugs/therapeutic-area/34/genetic-disease

156. Wu, Y. et al. Edição genética terapêutica altamente eficiente de células-tronco hematopoiéticas humanas. *Nat. Med* . **25** , 776–783 (2019).

157. Jaiswal, S. et al. Hematopoiese clonal e risco de doença cardiovascular aterosclerótica. *N. Engl. J. Med* . **377** , 111–121 (2017).

158. Lee, MH et al. Recombinação somática do gene APP na doença de Alzheimer e neurônios normais. *Natureza* **563** , 639–645 (2018).

159. Consórcio H3Africa. Permitir a revolução genómica em África. *Ciência* **344** , 1346–1348 (2014).

160. Dolman, L. et al. ClinGen avançando nos padrões de compartilhamento de dados genômicos como um projeto driver GA4GH. *Zumbir. Mutação* . **39** , 1686–1689 (2018).

161. Fairfax, BP et al. A atividade imune inata condiciona o efeito de variantes regulatórias na expressão gênica de monócitos. *Ciência* **343** , 1246949 (2014).

162. Nelson, MR et al. O apoio de evidências genéticas humanas para indicações de medicamentos aprovados. *Nat. Geneta* . **47** , 856–860 (2015). **Um artigo influente que demonstra que a probabilidade de sucesso terapêutico é aumentada (em cerca de um fator de 2) para alvos de medicamentos com evidências de apoio da genética humana.**

163. Veitch, DP et al. Compreender a progressão da doença e melhorar os ensaios clínicos da doença de Alzheimer: destaques recentes da Iniciativa de Neuroimagem da Doença de Alzheimer. *Demência de Alzheimer* . **15** , 106–152 (2019).

164. Ashley, EA Rumo à medicina de precisão. *Nat. Rev.* **17** , 507–522 (2016).

165. Franks, PW & McCarthy, MI Expondo as exposições responsáveis pelo diabetes tipo 2 e obesidade. *Ciência* **354** , 69–73 (2016).

166. Sanna, S. et al. Relações causais entre o microbioma intestinal, ácidos graxos de cadeia curta e doenças metabólicas. *Nat. Geneta* . **51** , 600–605 (2019).

167. Udler, MS et al. Os loci genéticos do diabetes tipo 2 informados por associações multicaracterísticas apontam para mecanismos e subtipos da doença: uma análise de agrupamento suave. *PLoS Med* . **15** , e1002654 (2018).

168. Rose, G. Indivíduos doentes e populações doentes. *Internacional J. Epidemiol* . **14** , 32–38 (1985).

169. Hindorff, LA et al. Priorizando a diversidade na pesquisa em genômica humana. *Nat. Rev.* **19** , 175–185 (2018).

170. Wittrup, A. & Lieberman, J. Derrubando a doença: um relatório de progresso sobre a terapêutica com siRNA. *Nat. Rev.* **16** , 543–552 (2015).

171. Schuster, SJ et al. Células T receptoras de antígeno quimérico em linfomas de células B refratários. *N. Engl. J. Med* . **377** , 2545–2554 (2017).

172. Ding, Q. et al. Alteração permanente de PCSK9 com edição in vivo do genoma CRISPR – Cas9. *Circ. Res* . **115** , 488–492 (2014).

173. Risca, VI & Greenleaf, WJ Desvendando o genoma 3D: ferramentas genômicas para exploração multiescala. *Tendências Genet* . **31** , 357–372 (2015).

Reconhecimentos

Reconhecemos o financiamento dos seguintes financiadores. JHC: NIH (U01DK062429, U01DK062422 e R01DK106593); NJC: NIH (U01HG009086, U54MD010722 e R01MH113362); ETD: Fundação Nacional Suíça para a Ciência, Fundação Louis Jeantet; EEK: NIH (R01HG010297, R01HL104608, U01HG009610, R01DK110113, U01HG009080, X01HL1345 e UM1HG0089001); LMC: NIHR Oxford Biomedical Research Center e NIH (5P50HD028138-27); KNN: NHMRC (APP1113531);

SET: NIH (U01HG006485 e 1U41HG009649); CNR: Programa Intramural do NIH no Centro de Pesquisa em Genômica e Saúde Global e Instituto Nacional de Pesquisa do Genoma Humano; e MIM: Wellcome (090532, 098381, 106130, 203141 e 212259) e NIH (U01-DK105535, DK085545 e DK098032). O financiamento pessoal vem das seguintes fontes. MC: Fundo de Próxima Geração do Broad Institute do MIT e Harvard; JHC: Sanford Grossman Charitable Trust e Helmsley Charitable Trust; RC: UKBiobank; LMC: Fundação Li Ka Shing; JS: Instituto Médico Howard Hughes; e MIM: Wellcome e NIHR.

Informação sobre o autor

Autores e Afiliações

Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, EUA
Melina Claussnitzer

Broad Institute do MIT e Harvard Cambridge, Cambridge, MA, EUA
Melina Claussnitzer, Sekar Kathiresan, Cecilia M. Lindgren, Daniel G. MacArthur e Heidi L. Rehm

Instituto de Ciências Nutricionais, Universidade de Hohenheim, Stuttgart, Alemanha
Melina Claussnitzer

Instituto Charles Bronfman de Medicina Personalizada, Icahn School of Medicine at Mount Sinai, Nova York, NY, EUA
Judy H. Cho e Eimear E. Kenny

Departamento de Medicina, Icahn School of Medicine at Mount Sinai, Nova York, NY, EUA
Judy H. Cho

Departamento de Genética e Ciências Genômicas, Icahn School of Medicine at Mount Sinai, Nova York, NY, EUA
Judy H. Cho e Eimear E. Kenny

Departamento de Saúde Populacional de Nuffield (NDPH), Universidade de Oxford, Oxford, Reino Unido
Rory Collins

Biobanco do Reino Unido, Stockport, Reino Unido
Rory Collins

Instituto de Genética Vanderbilt e Divisão de Medicina Genética, Departamento de Medicina, Centro Médico da Universidade Vanderbilt, Nashville, TN, EUA
Nancy J. Cox

Departamento de Medicina Genética e Desenvolvimento, Faculdade de Medicina da Universidade de Genebra, Genebra, Suíça
Emmanouil T. Dermitzakis

Centro do Genoma Health 2030, Genebra, Suíça
Emmanouil T. Dermitzakis

Instituto Wellcome Sanger, Hinxton, Reino Unido
Matthew E. Hurles e Nicole Soranzo

Centro de Medicina Genômica, Massachusetts General Hospital, Boston, MA, EUA
Sekar Kathiresan, Daniel G. MacArthur e Heidi L. Rehm

Verve Therapeutics, Cambridge, MA, EUA
Sekar Kathiresan

Centro de Saúde Genômica, Icahn School of Medicine at Mount Sinai, Nova York, NY, EUA
Eimear E. Kenny

Instituto de Big Data no Centro Li Ka Shing para Informação e Descoberta em Saúde, Universidade de Oxford, Oxford, Reino Unido
Cecília M. Lindgren

Centro Wellcome de Genética Humana, Departamento de Medicina de Nuffield, Universidade de Oxford, Oxford, Reino Unido
Cecilia M. Lindgren e Mark I. McCarthy

Unidade de Genética Analítica e Translacional, Massachusetts General Hospital, Boston, MA, EUA
Daniel G. MacArthur e Heidi L. Rehm

Instituto de Pesquisa Infantil Murdoch, Parkville, Victoria, Austrália
Kathryn N. Norte

Universidade de Melbourne, Parkville, Victoria, Austrália
Kathryn N. Norte

Departamentos de Pediatria e Genética Molecular e Humana, Baylor College of Medicine, Houston, TX, EUA
Sharon E. Plon

Centro de Câncer Infantil do Texas, Hospital Infantil do Texas, Houston, TX, EUA
Sharon E. Plon

Departamento de Patologia, Harvard Medical School, Boston, MA, EUA
Heidi L.Rehm

Instituto de Genética Humana, Universidade da Califórnia, São Francisco, São Francisco, CA, EUA
Neil Risch

Centro de Pesquisa em Genômica e Saúde Global, Instituto Nacional de Pesquisa do Genoma Humano, Bethesda, MD, EUA
Charles N. Rotimi

Departamento de Ciências do Genoma, Universidade de Washington, Seattle, WA, EUA
Jay Shendure

Instituto Brotman Baty de Medicina de Precisão, Magnuson Health Sciences Building, Seattle, WA, EUA
Jay Shendure

Instituto Médico Howard Hughes, Seattle, WA, EUA
Jay Shendure

Departamento de Hematologia, Universidade de Cambridge, Cambridge, Reino Unido
Nicole Soranzo

Centro de Oxford para Diabetes, Endocrinologia e Metabolismo, Oxford, Reino Unido
Mark I. McCarthy

Centro de Pesquisa Biomédica Oxford NIHR, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, Reino Unido
Mark I. McCarthy

Genética Humana, Genentech, South San Francisco, CA, EUA
Mark I. McCarthy

Contribuições

MC e MIM coordenaram a elaboração da Revisão. Todos os demais autores contribuíram na redação do texto e aprovaram a versão final.

autor correspondente

Correspondência para Mark I. McCarthy .

Declarações éticas

Interesses competitivos

RC recebeu bolsas de pesquisa da British Heart Foundation, Cancer Research UK, Medical Research Council, Merck & Co, UKBiobank, Wellcome e Medco; recebeu o Prêmio Pfizer (para o NDPH) e foi nomeado em uma patente para um teste genético de miopatia relacionado a estatinas. A RC não recebe nenhuma remuneração pessoal destes: qualquer parcela de royalties ou outros pagamentos foi dispensada em favor do NDPH. ETD é presidente e membro do conselho da Hybridstat e faz parte do conselho consultivo da DNAnexus. MEH é cofundador, acionista e diretor da Congenica. SK recebeu bolsas de pesquisa da Bayer e da Novartis; faz parte dos conselhos consultivos científicos do Regeneron Genetics Center, Corvidia Therapeutics e Maze Therapeutics; possui participação acionária na San Therapeutics, Catabasis, Verve e Maze Therapeutics e é consultor da Maze Therapeutics, Alynlam, ExpertConnect, Leerink Partners, Noble Insights, Bayer e Novo Ventures. EEK recebe honorários da Illumina e Regeneron Pharmaceuticals. A CML mantém colaborações de pesquisa com a Novo Nordisk e a Bayer, não recebendo nenhum pagamento pessoal. A DGM é cofundadora e acionista da Goldfinch Bio e recebeu financiamento de pesquisa da AbbVie, Biogen, BioMarin, Merck, Pfizer e Sanofi-Genzyme. SEP é membro do Conselho Consultivo Científico da Baylor Genetics. JS é membro do Conselho Consultivo Científico da Maze Therapeutics, Camp4 Therapeutics, Nanostring, Phase Genomics, Adaptive Biotechnology e Stratos Genomics, fundador da Phase Genomics e consultor da Guardant Health. MIM foi membro de painéis consultivos da Pfizer, NovoNordisk e Zoe Global; recebeu honorários da Merck, Pfizer, NovoNordisk e Eli Lilly e financiamento de pesquisa da Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly,

Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier e Takeda. Em junho de 2019, MIM é funcionário da Genentech e detentor de ações da Roche. As opiniões expressas neste artigo são de responsabilidade do(s) autor(es) e não necessariamente do NHS, do NIHR ou do Departamento de Saúde.

Informações adicionais

Informações sobre revisão por pares *A Nature* agradece a Ewan Birney, Jeff Schloss, Richard Trembath e ao(s) outro(s) revisor(es) anônimo(s) por sua contribuição para a revisão por pares deste trabalho.

Nota do editor A Springer Nature permanece neutra em relação a reivindicações jurisdicionais em mapas publicados e afiliações institucionais.

Direitos e permissões

Reimpressões e permissões

Sobre este artigo

Cite este artigo

Claussnitzer, M., Cho, JH, Collins, R. *et al.* Uma breve história da genética das doenças humanas. *Natureza* **577** , 179–189 (2020). https://doi.org/10.1038/s41586-019-1879-7

Recebido	Aceitaram	Publicados
16 de julho de 2019	13 de novembro de 2019	08 de janeiro de 2020

Data de emissão

09 de janeiro de 2020

DOI

https://doi.org/10.1038/s41586-019-1879-7

Compartilhe este artigo

Qualquer pessoa com quem você compartilhar o link a seguir poderá ler este conteúdo:

Obtenha link compartilhável

Fornecido pela iniciativa de compartilhamento de conteúdo Springer Nature SharedIt

assuntos Genômica funcional • Genética médica

Este artigo é citado por

Efeitos genéticos de elementos semelhantes a intensificadores conservados em sequência em características complexas humanas

Xiang Zhu, Mãe brilhante & Asa Pendurada Wong

Biologia do Genoma (2024)

A interação da ética e das tecnologias genéticas no equilíbrio da valoração social do genoma humano nas declarações da UNESCO

Christina Gaydarska, Kayo Takashima ... Jusaku Minari

Jornal Europeu de Genética Humana (2024)

Cinco etapas para conectar variantes de risco genético a doenças

Natureza (2024)

A tela massivamente paralela revela muitas variantes raras de 3' UTR que regulam a abundância de mRNA de genes causadores de câncer

Ting Fu, Kofi Amoah ... Xin Shu Xiao

Comunicações da Natureza (2024)

Autoencodificador variacional para projeto de sorotipos de vetores virais sintéticos

Suyue Lyu, Shahin Sowlati-Hashjin & Michael Garton

Inteligência da Máquina da Natureza (2024)

Natureza (*Natureza*) | ISSN 1476-4687 (on-line) | ISSN 0028-0836 (imprimir)